Structural studies of a thermostable citrate synthase

McCormack, Michelle

Award date:
1995

Awarding institution:
University of Bath

STRUCTURAL STUDIES OF A THERMOSTABLE CITRATE SYNTHASE

submitted by Michelle M^{c}Cormack

for the degree of PhD

of the University of Bath

1995

Copyright

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisors Mike Danson, David Hough and Garry Taylor for their advice, encouragement and friendship during the past three years. I believe that I was one of the most fortunate members of the Biochemistry Dept., in having them as my supervisors. I would also like to thank Zeneca BioProducts, and in particular David Byrom, for industrial sponsorship.

Thanks go to members of the crystallographic unit, Susan, Rupert and Andrew for all their assistance with the homology-modelling and never complaining when I disturbed their peace ! I also wish to thank all the members (past and present) of the 'End-lab' for their invaluable help and making life Jolley !

Special thanks go to Gordon Munro (lab oracle and councillor rolled into one), and to Dibble (essential in times of kinetic crisis), for being such wonderful friends.

Finally, I will never be able to thank this person enough for all his enthusiastic help with many aspects of this thesis, and for his encouragement, patience, understanding and love. Thanks, Jim.

ABSTRACT

The work presented in this thesis complements that of our research group, investigating the thermal stability, halophilicity and psychrophily of a range of archaeal enzymes. Citrate synthase has been chosen as a model to investigate protein thermostability.

1) The 3D structure of citrate synthase from *Thermoplasma acidophilum*, a moderately thermophilic Archaeon, was homology modelled. The model was then compared to the later-determined crystal structure of the enzyme to investigate the validity of the modelling technique. The results of the analysis suggest that the most sensitive areas of the procedure are the multiple sequence alignment and the manual technique employed.

2) The structures (primary to quaternary) of the citrate synthases from pig and *Tp. acidophilum* were used in a comparative analysis in order to identify features that may play a role in conferring thermal stability upon the archaeal enzyme. The trend from mesophile to thermophile, appears to be towards a more compact and potentially less flexible molecule, achieved in the thermostable protein through additional interactions, fewer cavities, increased α-helix stability, better complementarity between the subunits of the dimer and an absence of flexible regions, such as loops and surface helices. The role of three of the features was investigated by site-directed mutagenesis, and in each case pig citrate synthase was mutated to possess the archaeal properties.

(a) The level of hydrophobic interactions at the subunit interface was increased, resulting in mutant proteins that were slightly less thermostable, and with an altered catalytic efficiency. These differences appear to be due to small conformational changes within the secondary and tertiary structures of the mutants.

(b) Cavity-filling mutants and a truncated mutant of pig citrate synthase were also created. In both cases, the proteins were expressed in *E. coli* in inactive forms.

These site-directed mutagenesis results did not lead to the identification of positive stabilising features in citrate synthase and thus alternative pathways for investigating thermostability in the enzyme are also discussed.

ABBREVIATIONS

AMINO ACIDS (SINGLE AND TRIPLE LETTER CODES)

Ala	(A)	Alanine	Leu	(L)	Leucine
Arg	(R)	Arginine	Lys	(K)	Lysine
Asn	(N)	Asparagine	Met	(M)	Methionine
Asp	(D)	Aspartic acid	Phe	(F)	Phenylalanine
Cys	(C)	Cysteine	Pro	(P)	Proline
Gln	(Q)	Glutamine	Ser	(S)	Serine
Glu	(E)	Glutamic acid	Thr	(T)	Threonine
Gly	(G)	Glycine	Trp	(W)	Tryptophan
His	(H)	Histidine	Tyr	(Y)	Tyrosine
Ile	(I)	Isoleucine	Val	(V)	Valine

AcCoA	Acetyl coenzyme A
AFOR	Aldehyde ferredoxin oxidoreductase
ATP	Adenosine 5'-triphosphate
ATPase	Adenosine triphosphatase
BSA	Bovine serum albumin
C_p	Heat capacity
CD	Circular dichroism
β–CD	β–Cyclodextrin
CoA	Coenzyme A
CS	Citrate synthase
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease
DSC	Differential scanning calorimetry
DTNB	5,5'-dithio-bis(2-nitrobenzoic acid)
DTT	Dithiothreitol
EDTA	Ethylenediaminetetra-acetic acid
FAD(H_2)	Flavin adenine dinucleotide
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
IPMDH	Isopropyl malate dehydrogenase
IPTG	Isopropyl-β-D-thiogalactoside
MDH	Malate dehydrogenase
MQW	MilliQ water

NAD(H)	Nicotinic acid adenine dinucleotide
NMR	Nuclear magnetic resonance
NTP	Nucleoside 5'-triphosphate
dNTP	Deoxynucleoside 5'-triphosphate
OAA	Oxaloacetic acid
P_i	Ortho phosphate ion
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PGK	Phosphoglycerate kinase
POE(10)L	Polyoxyethylene 10 lauryl
RMS	Root mean square
RNA	Ribonucleic acid
tRNA	Transfer RNA
rRNA	Ribosomal RNA
RNAse	Ribonuclease
SDS	Sodium dodecyl sulphate
SDS-PAGE	SDS-Polyacrylamide gel electrophoresis
T_m	Melting temperature
TAE	40 mM Tris.acetate, 1 mM EDTA
TBE	90 mM Tris.borate, 2 mM EDTA
TCA	Trichloroacetic acid
TE8	10 mM Tris.HCl, 2 mM EDTA, pH8.0
2TE8	20 mM Tris. Hcl, 2 mM EDTA, pH8.0
TEMED	N,N,N',N'-tetramethylethylenediamine
Tris	Tris-(hydroxymethyl)-methylamine
X-Gal	5-Bromo-4-chloro-3-indolyl-β-D-galactoside

CONTENTS

CHAPTER 1

INTRODUCTION

1.1 CITRATE SYNTHASE

The work presented in this thesis aims to contribute to our understanding of the structural basis of protein thermostability. We have chosen the Archaea as a source of hyperstable proteins to study and have selected one enzyme, citrate synthase, to analyse in detail. Therefore, the introduction to this thesis concerns a detailed description of the chosen enzyme, citrate synthase, a brief account of the Archaea, and finally a consideration of what is currently known about protein thermostability.

1.1.1 COMPARATIVE ANALYSES OF CITRATE SYNTHASE

Citrate synthase is a central metabolic enzyme found in virtually all known organisms. It is the first enzyme of the citric acid cycle (Fig. 1.1), controlling the entry of carbon units into the cycle [Krebs & Lowenstein, 1960]. The reaction catalysed by citrate synthase involves the condensation between oxaloacetate and acetyl coenzyme A to form citryl coenzyme A, with subsequent hydrolysis to form citrate and coenzyme A.

Citrate synthases from the domains Eucarya and Bacteria have been compared. Studies reveal a correlation between the taxonomic status of the organism and the subunit composition, catalytic activity and regulatory properties of the enzyme [Weitzman & Danson, 1976; Weitzman, 1981; Danson, 1988]. Those from Eucarya and Gram-positive Bacteria exist as dimers that are isosterically inhibited by ATP, whilst the citrate synthases from Gram-negative bacteria are hexameric and are

allosterically inhibited by NADH, showing no regulatory response to ATP. The exceptions to this rule are the dimeric citrate synthases from the Gram-negative organisms *Coxiella burnetii*, *Rickettsia prowazekii* and *Thermus aquaticus* that are regulated by ATP also.

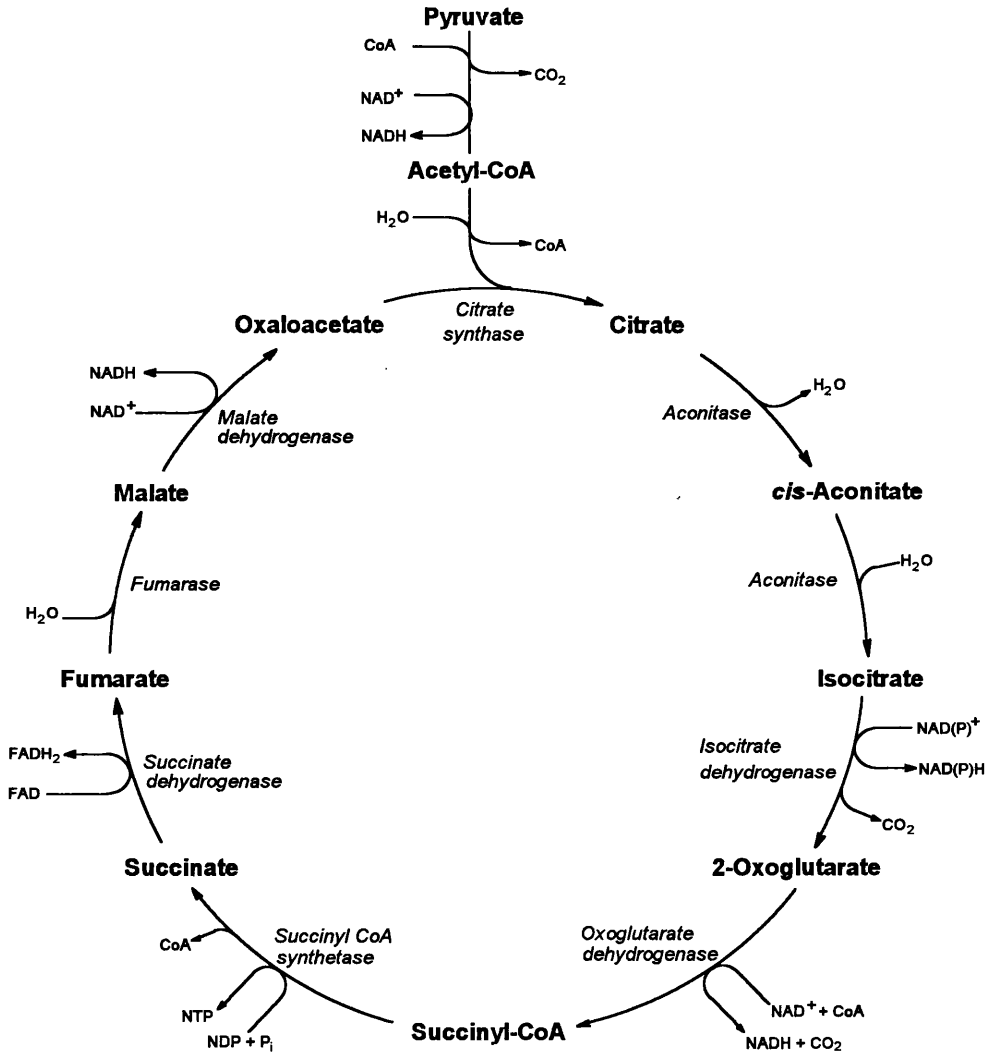


Fig. 1.1 Citric acid cycle illustrating the position of citrate synthase.

The presence of citrate synthase in all three archaeal phenotypes has been demonstrated [Danson *et al.*, 1985]. It has been shown that the enzymes from the Archaea *Sulfolobus acidocaldarius*, *Thermoplasma acidophilum*, *Haloferax volcanii* and *Pyrococcus furiosus* are dimeric and possess similar regulatory and catalytic properties to the dimeric citrate synthases from Eucarya and Gram-positive Bacteria [Smith *et al.*, 1987; Lohlein-Werhahn *et al.*, 1988; Grossebuter & Gorrisch, 1985; James *et al.*, 1991; Muir *et al,*. 1995].

All citrate synthases comprise identical subunits with relative molecular masses of approximately 45,000 Da. The hexamer is thought to function as a trimer of dimers [Else *et al.*, 1988] and thus it would seem that the oligomeric status influences the regulatory properties of the enzyme. A second citrate synthase gene has been identified in *E. coli* [Patton *et al.*, 1993] and *Pseudomonas aeruginosa* [Anderson *et al.,* 1993]. Both of these additional citrate synthases are truncated and, from the limited sequence information of their N-terminal peptides, they appear to align more closely with the *Tp. acidophilum* citrate synthase sequence than they do to their own hexameric citrate synthase.

1.1.2 STRUCTURE OF CITRATE SYNTHASE

There is an expanding database of structural information concerning citrate synthase. Amino acid sequences of the enzyme are available from 20 different sources, with representatives from the Eucarya, Bacteria and the Archaea (Table 1.1). Two of the sequences are from archaeal sources, the moderate thermophile *Tp. acidophilum* and the hyperthermophile *P. furiosus*, and three of the proteins are thermostable, that is the two archaeal enzymes, and a third from the thermotolerant *Bacillus coagulans* (strain C4). There are also two archaeal citrate synthase N-terminal peptide sequences, derived from the halophile *H. volcanii*, and the extreme thermophile *S. solfataricus*.

SOURCE ORGANISM	DOMAIN	REFERENCE
Pig	Eucarya	Bloxham *et al.*, 1981, 1982
Chicken		Karpusas *et al.*,1990
Saccharomyces cerevisiae		
glyoxysomal		Rosenkrantz *et al.*, 1986
mitochondrial		Suissa *et al.*, 1984
Arabidopsis thaliana		Unger *et al.*, 1989
Tetrahymena thermophila		Numato *et al.*, 1991
Caenorhabditis elegans		Swiss-Prot Data Bank - P34575
Neurospora crassa		EMBL Data Bank - P34085
Bacillus coagulans (strain C4)	Bacteria (Gram + ve)	Schendel *et al.*, 1992
Bacillus subtilis (CitZ)		Jin and Sonenshein, 1994
Bacillus subtilis (CitA)		Jin and Sonenshein, 1994
Mycobacterium smegmatis		David *et al.*, 1991
Pseudomonas aeruginosa	Bacteria (Gram - ve)	Donald *et al.*, 1989
Acinetobacter anitratum		Donald & Duckworth, 1986
Escherichia coli		Ner *et al.*, 1983
Acetobacter acetii		Fukaya *et al.*, 1990
Coxiella burnetii		Heinzen *et al.*, 1991
Rickettsia prowazekii		Wood *et al.*, 1987
Thermoplasma acidophilum	Archaea	Sutherland *et al.*, 1990
Pyrococcus furiosus		Muir *et al.*, 1995
Haloferax volcanii	Archaea (N-terminal peptide)	James *et al.*, 1991
Sulfolobus solfataricus		Lill *et al.*, 1992

Table 1.1 Sources of citrate synthases for which amino acid sequences are available. The Data Bank Accession numbers are shown for those that are not published sequences.

The crystal structures of citrate synthase from pig heart and chicken heart have been solved by multiple isomorphous replacement to 2.0 Å and 1.7 Å, respectively [Remington *et al.*, 1982]. There are two crystal forms of the enzyme, the tetragonal 'open' form in the absence of ligands, and the monoclinic 'closed' form which was crystallised with citrate and coenzyme A bound. The crystal structure of the open form of *Tp. acidophilum* citrate synthase has also been solved to 2.5 Å by molecular

replacement [Russell *et al.*, 1994]. The crystal structure of a hexameric citrate synthase has not yet been determined, although the protein from *E. coli* has been crystallised [Rubin *et al.*, 1983].

The crystal structure of pig citrate synthase reveals that 315 of the 437 residues (72%) form α-helices; these are labelled A to T and are shown below (Table 1.2). There are also two small β-strands. The 3D structure of *Tp. acidophilum* citrate synthase is very similar to that from pig, with an overall rms deviation of 2.27 Å for the 356 α-carbon atoms that have been refined. The structure of the archaeal citrate synthase does not include the initial 6 N-terminal residues, nor the final 14 C-terminal residues, due to poorly-defined electron density in these areas. The α–helical content of the archaeal structure is 57% of all the residues refined. The 16 helices of the archaeal structure are labelled using the same nomenclature as for pig citrate synthase and are shown in Table 1.2.

The subunit interface of the citrate synthase dimer is made up of a four helices (F, G, M and L) that are anti-parallel to the same four of the other subunit. Upon dimer formation a region of the monomer surface becomes inaccessible to solvent, approximating to 28% and 20% for the pig and archaeal citrate synthases, respectively. In pig citrate synthase the majority of the residues for which solvent accessibility decreases by more than 5 $Å^2$ are hydrophobic, and almost every residue that is hydrophilic is involved in intersubunit hydrogen bonds and salt bridges. In both structures, wrapped around the interface helical sandwich is a pair of anti-parallel helices, I and S, which both bend smoothly. Helix I of the pig enzyme contains a Pro residue that is thought to induce the bending, whilst *Tp. acidophilum* citrate synthase contains a serine at the equivalent position. Helices A, P and R in the pig enzyme structure are also kinked, the former two due to intrahelical Pro residues.

PIG CS						*TP. ACIDOPHILUM* CS					
A	Asn	5	-	Gly	29						
B	Thr	37	-	Gly	43						
C	Ser	70	-	Leu	78	**C**	Val	37	-	Ala	43
D	Leu	88	-	Gly	99	**D**	Asp	47	-	Tyr	56
E	Thr	103	-	Ala	118	**E**	Glu	62	-	Val	74
F	Pro	121	-	Phe	131	**F**	Asp	80	-	Arg	87
G	His	136	-	Ser	152	**G**	Ala	95	-	Ser	109
H	Asn	153	-	Gly	161				-		
I	His	163	-	Arg	195	**I**	Asp	120	-	Ile	142
J	Asp	208	-	Gly	218	**J**	Tyr	156	-	Phe	165
K	Asp	221	-	Ser	236	**K**	Lys	171	-	Tyr	183
L	Asn	242	-	Leu	255	**L**	Ala	191	-	Val	200
M	Asp	257	-	Gly	271	**M**	Asp	206	-	Leu	217
N	His	274	-	Glu	291	**N**	Ala	225	-	Ile	236
O	Ser	297	-	Gly	312	**O**	Val	242	-	Asn	249
P	Asp	327	-	Leu	341	**P**	Pro	270	-	Ser	283
Q	Asp	344	-	Gly	365	**Q**	Lys	287	-	Phe	307
R	Asn	373	-	Gly	386	**R**	Thr	317	-	Ile	327
S	Met	390	-	Gly	416	**S**	Ile	335	-	Glu	359
T	Ser	426	-	Leu	433						
1	Val	57	-	Pro	60	**1**	Thr	21	-	Asp	24
2	Glu	62	-	Arg	65	**2**	Ile	29	-	Tyr	32
			-			**3**	Tyr	35	-	Ser	36

Table 1.2 Residues involved in the secondary structural elements of citrate synthase (CS) from pig and *Tp. acidophilum*. The nomenclature for pig citrate synthase is used for both enzymes, where the α-helices are labelled A to T and the β-strands are labelled 1 to 3.

1.1.3 CATALYTIC MECHANISM OF CITRATE SYNTHASE

The monomer of citrate synthase can be divided into two domains, a small domain consisting of helices N, O, P, Q and R, and a large domain. The orientation of the domains is significant for citrate synthase catalysis. From binding studies and Michaelis-Menten rate behaviour there is evidence for an ordered reaction mechanism, with oxaloacetate binding first [Johansson & Pettersson, 1974]. The affinity for acetyl coenzyme A increases at least 20-fold when oxaloacetate is previously bound, indicating strong heterotropic cooperativity [Johansson & Pettersson, 1977]. From crystallographic studies of the substrate-bound pig enzyme, substrate binding has been shown to coincide with a change in protein conformation [Remington *et al.*, 1982], the small domain rotating 18.5° relative to the large domain, thus enclosing the oxaloacetate in a 'pocket' whilst forming a binding-site for acetyl coenzyme A. The closed form is more compact than the open form with a decrease of approximately 7% of solvent-accessible surface area upon transition from open to closed forms. This compactness of the molecule may be a factor that explains the increased stability to denaturation of citrate synthase bound with oxaloacetate or inhibitors [Srere, 1963].

The catalytic mechanism has been studied extensively in pig heart citrate synthase through work on crystal structures of the enzyme bound to substrates, substrate analogs, transition state analogues and products [Karpusas *et al.*, 1990, 1991; Remington, 1992], and also by site directed mutagenesis of active site residues [Alter *et al.*, 1990; Zhi *et al.*, 1991]. The residues involved at the active site of the pig enzyme, and also the putative residues in *Tp. acidophilum* citrate synthase are shown below (Table 1.3). The archaeal residues have been identified by docking the substrates into the crystal structure of *Tp. acidophilum* citrate synthase to indicate possible interactions. There are two independent active sites per dimer with residues from both subunits contributing to each. The C-terminal arm of each monomer lies within a groove in the opposite subunit and provides catalytic residues (Fig. 1.2).

PIG CS	*TP. ACIDOPHILUM* CS
Arg 46	**Arg 364**
Arg 164	**Arg 366**
His 238	His 187
His 274	His 222
His 320	His 262
Arg 324	Lys 266
Arg 329	Arg 271
Lys 368	
Asp 375	Asp 317
Arg 401	Arg 344
Arg 421	**Arg 361**

Table 1.3 Active site residues of pig citrate synthase [Remington *et al.*, 1982] and the analogous residues in *Tp. acidophilum* citrate synthase that perform the equivalent interactions to the pig enzyme residues. Those amino acids in bold are contributions from the second subunit in the dimeric molecule. The catalytic residues are shown in italics.

Remington [1992] has proposed that the reaction mechanism of pig citrate synthase progresses in three stages (Fig. 1.3). The first step is the enolisation of acetyl coenzyme A involving a concerted acid-base mechanism. His 274 protonates the carbonyl oxygen of the acetyl coenzyme A, whilst Asp 375 simultaneously deprotonates the methyl group of the substrate. This step is followed by the condensation of the enol with oxaloacetate, using a second acid-base mechanism. His 320 protonates the carbonyl oxygen of oxaloacetate, whilst His 274 is reprotonated in preparation for the next catalytic cycle. The carbanionic methyl group of the acetyl coenzyme A nucleophilically attacks the activated carbonyl of oxaloacetate to form the intermediate citryl coenzyme A. The final step is the hydrolysis of this intermediate at the thioester bond to release citrate and coenzyme A. This step is not fully understood but two mechanisms have been proposed. Remington [1992] suggests that Asp 375 protonates the sulphur of acetyl coenzyme A which is then susceptible to hydrolysis by an activated water molecule (Fig. 1.3(a)). Alter *et al.* [1990] propose that Asp 375 and the intermediate form an anhydride complex, releasing coenzyme A, but remaining susceptible to attack by an activated water molecule to release citrate (Fig. 1.3(b)).

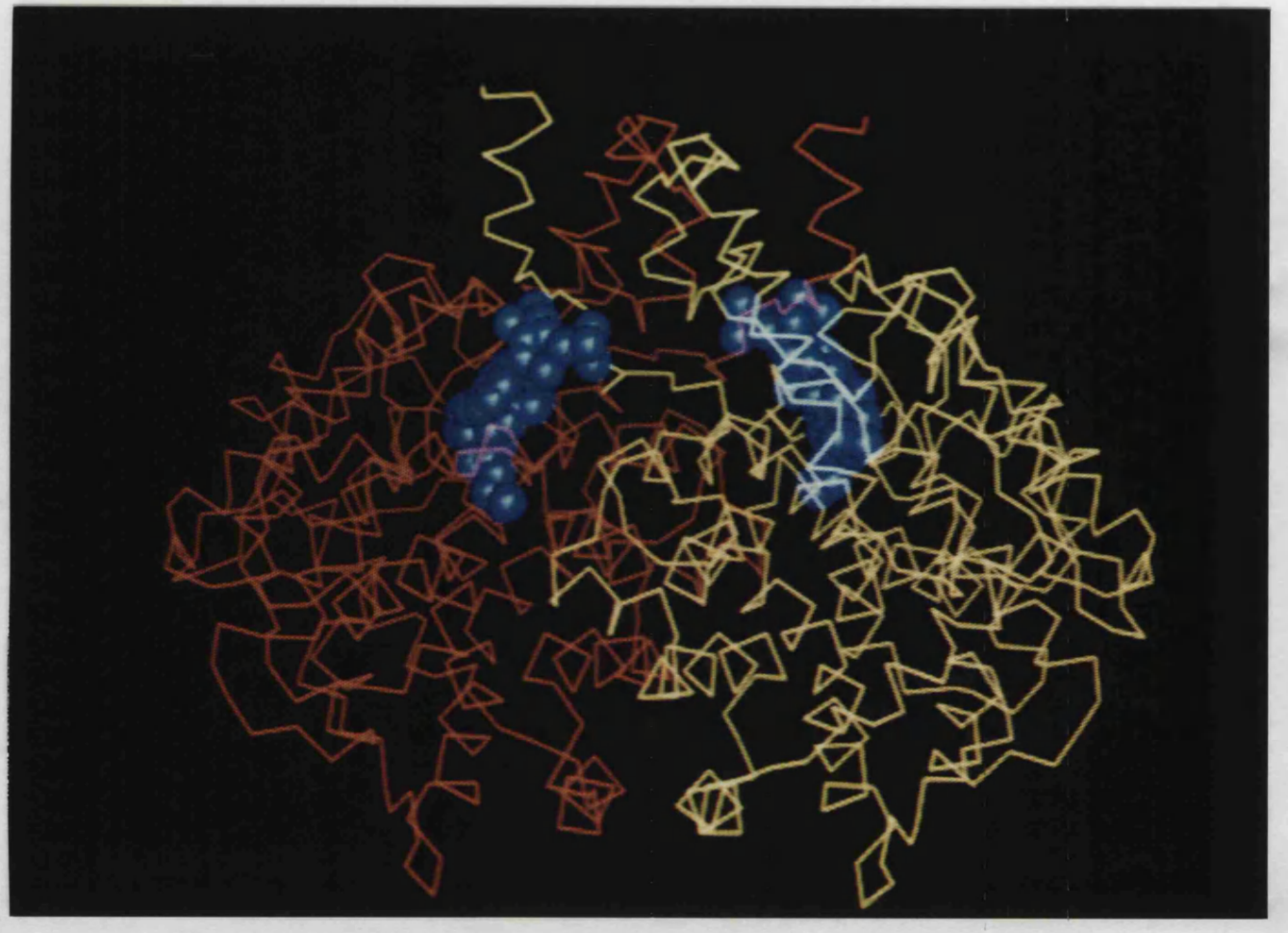


Fig. 1.2 The α-carbon trace of the closed form of the pig citrate synthase dimer (subunits are coloured red or yellow). The substrates (blue) are shown also. Note how the C-terminal arm of one subunit is positioned in a groove formed by the second subunit, and thus is able to contribute catalytic residues to the active site of the second subunit.

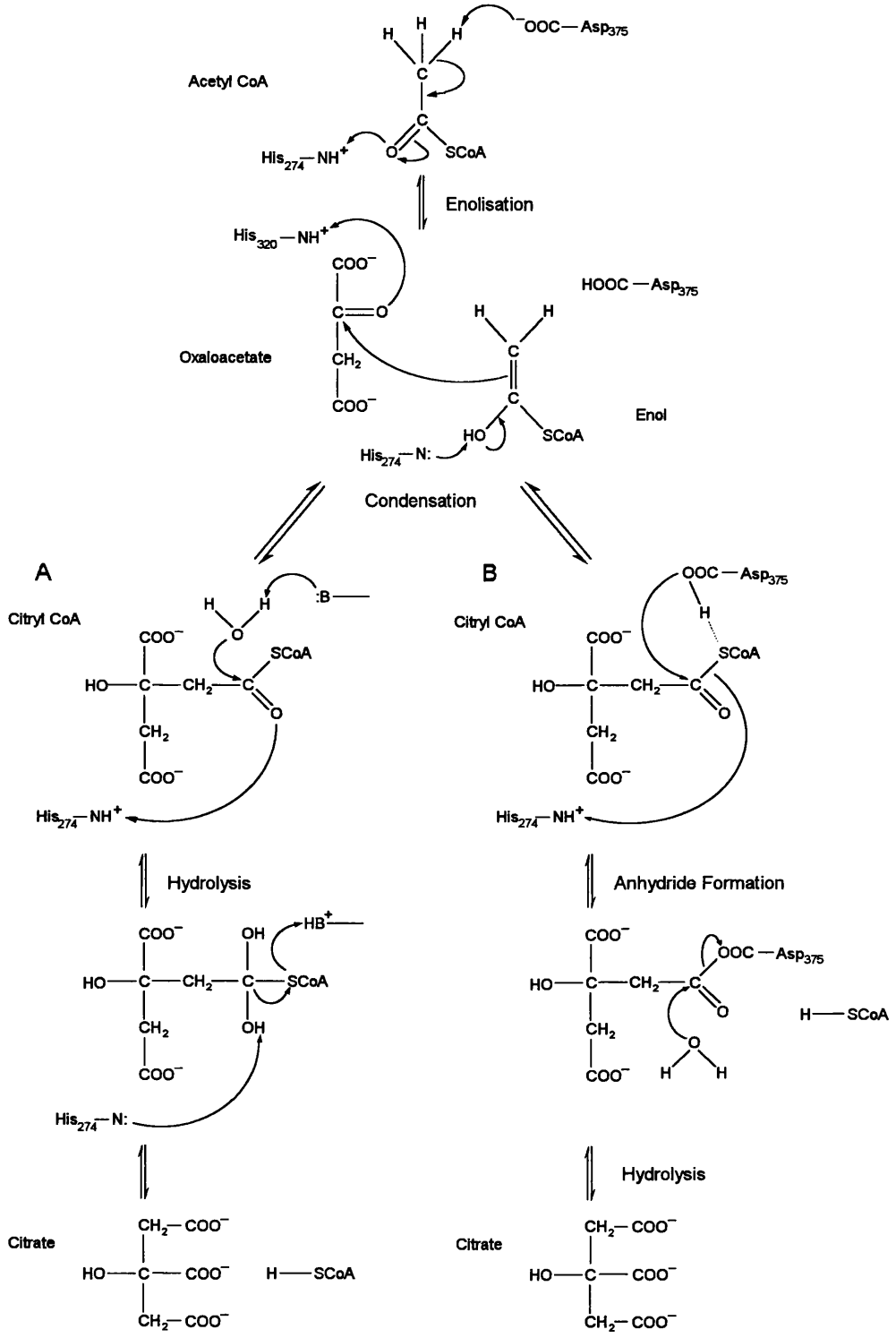


Fig. 1.3 The Remington model for the catalytic mechanism of pig citrate synthase [Remington, 1992]. There are three steps:

1. Enolisation of acetyl coenzyme A
2. Condensation with oxaloacetate
3. Hydrolysis to release citrate and coenzyme A.

There are two proposed mechanisms for the hydrolysis step proposed by Remington [1992] (A), and Alter *et al.*, [1990] (B).

1.2 THE ARCHAEA

1.2.1 ARCHAEAL PHYLOGENY

It was once said that the genome of an organism is the ultimate record of its evolutionary history. The rapid progress in molecular biology techniques witnessed in the last couple of decades has enabled a part of that history to be deciphered. Until 1977, all organisms were classified as belonging to one of two kingdoms, the Eukaryota and the Prokaryota. This classification was particularly unsound, foremost since it was initially based upon on negative facts, such that Prokaryotes were classified by their lack of morphological characteristics. Then, in 1977, from extensive molecular characterisation of 16S and 18S rRNA molecules from a variety of organisms Carl Woese and colleagues proposed that all living organisms could be divided into three taxonomic groups arising from individual lines of descent [Woese & Fox, 1977]. The Archaebacteria, as they were termed, were a newly-recognised group in their proposal that were phylogenetically distinct from the Eucaryotes and the Eubacteria. Today all organisms are classified as either Eucarya, Bacteria or Archaea [Woese *et al.*, 1990], the term bacteria having been dropped from Archaebacteria to denote their truly separate nature from the other prokaryotes. Additional theories on the phylogenetic relationships between organisms are still emerging using criteria other than rRNA sequences; however, Woese's theory of universal phylogeny [Woese & Olsen, 1986; Woese *et al.*, 1990] is the most widely accepted proposal at the present.

The comparisons of whole 16S and 18S rRNA sequences from organisms representing all three domains [referenced in Woese & Olsen, 1986] were used for the proposal of universal phylogeny. The methods of analysis used to construct the phylogenetic tree were distance matrix and maximum parsimony algorithms. Initially, only partial rRNA sequences were available and so only the three major subgroups of the Archaea were defined [Woese *et al.*, 1978], these being the halophiles, the methanogens and the thermophiles. As whole rRNA sequences

were determined then the branching order within the Archaea could be defined [Woese & Olsen, 1986] (Fig. 1.4). This tree is supported by well-defined archaeal phenotypic data [for review see Woese and Wolfe, 1985; Kates *et al.*, 1993].

The phylogenetic tree proposed by Woese from 16S rRNA sequences was rooted using data from two sets of duplicated genes selected by Iwabe and co-workers [Iwabe *et al.*, 1989]. The genes encoding the α and β subunits of F_1 ATPases and elongation factors EF-1a and EF-2 each duplicated before the proposed divergence of the progenote into the three domains of life. The positioning of the root of the tree has been supported by additional data from amino-acyl tRNA synthetase protein sequences [Brown & Doolittle, 1995].

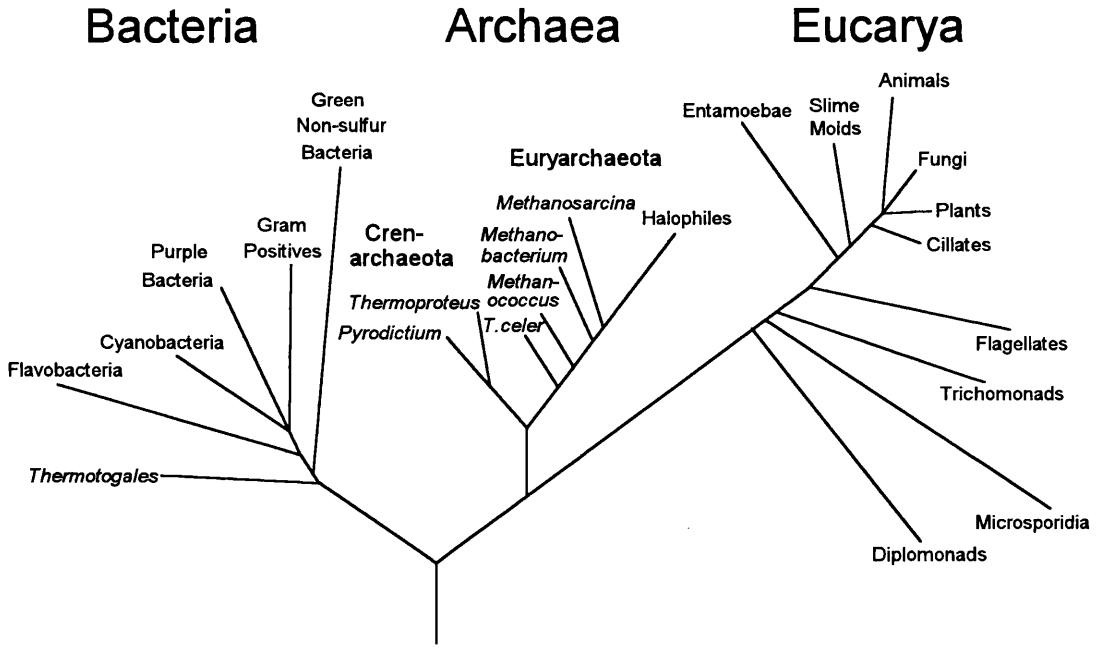


Fig. 1.4 (a) The universal rooted phylogenetic tree proposed by Woese et al., [1990] from 16S and 18S rRNA sequence comparisons using distance matrix and maximum parsimony algorithms. The tree was rooted using data from the duplicated genes encoding the α and β subunits of F_1 ATPases and elongation factors EF-1a and EF-2 [Iwabe *et al.*, 1989].

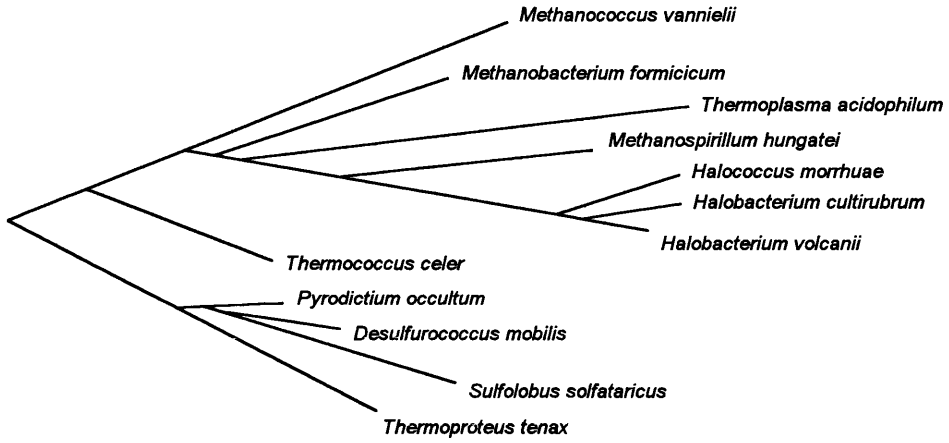


Fig 1.4 (b) Phylogenetic tree showing the branchings within the domain Archaea [Woese & Olsen, 1986]

Any molecule chosen as a molecular chronometer must be fundamental and essential to all living organisms since it is used to infer their evolutionary history. The 16S and 18S rRNA molecules, used by Woese, are ideal molecular chronometers with which to measure phylogenetic relationships since they are functionally conserved, ubiquitous to all organisms, easily isolated, and are a composite of both highly-conserved regions and more varied regions, allowing the phylogenetic relationship between both distantly related and closely related species to be measured. Other molecules and methods of analysis have been used to construct phylogenetic trees; similar algorithms to that of Woese have been used to analyse 5S rRNA secondary structural features [Fox *et al.*, 1982; Wolters & Erdmann, 1986], the structure of DNA-dependent RNA polymerases [Zillig *et al.*, 1985, 1989], ATPase gene sequences [Gogarten *et al.*, 1989], glutamate dehydrogenase gene sequences [Forterre *et al.*, 1993; Benachenhou-Lahfa *et al.*, 1993], and 70 kDa heat shock protein sequences [Gupta & Singh, 1994]. In addition, a phylogenetic tree has been proposed from the comparison of 16S rRNA sequences using an evolutionary parsimony algorithm [reviewed in Lake, 1991].

The two additional phylogenetic trees that have generated most discussion, and critical reviews from Woese [Woese & Olsen, 1986; Woese, 1987; Olsen & Woese, 1989], are those proposed by Zillig *et al.* [1989] and Lake [1991] (Fig. 1.5). From

the structural comparisons of DNA-dependent RNA polymerases, Zillig proposes that the Eucarya arose as the result of a fusion between the Bacteria and the Archaea, and supports his theory with data describing the shared characteristics between the Eucarya and Archaea, and the Eucarya and Bacteria. The chimeric origins of the Eucarya have also been suggested by Gupta's group [Gupta & Singh, 1994; Golding & Gupta, 1995]. Lake [1991], using a different part of the 16S rRNA sequence to Woese, employed an algorithm that claims to take into account the unequal-rate effect, i.e. the differing rates of evolution of either a protein or a genetic marker. Lake proposes that the Archaea are polyphyletic with the halophiles and methanogens grouped with the Bacteria and the thermophiles (named the 'eocytes') grouped with the Eucarya.

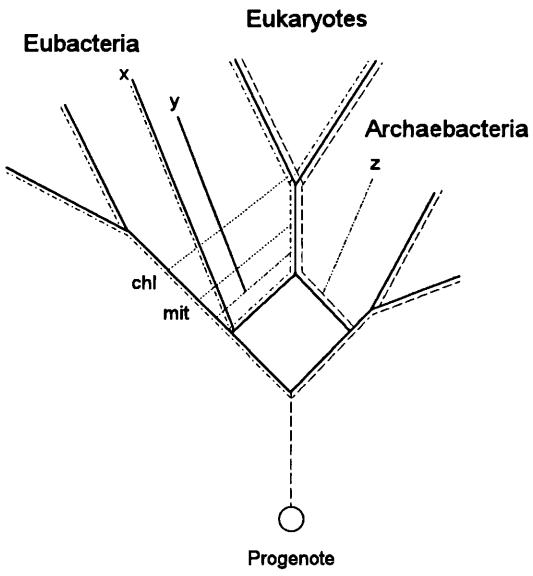


Fig. 1.5 (a) Schematic representation of the fusion hypothesis for the generation of the Eucaryotes, as proposed by Zillig *et al.* [1989] from structural comparisons of DNA-dependent RNA polymerases, using a maximum parsimony algorithm. The dotted lines indicate the acquisition of mitochondria and chloroplasts. The intermediary lines X, Y and Z indicate how for certain molecules certain lineages could branch from the eucaryotic lineage. Molecules inherited from a eubacterial ancestor (eg. pol 1) and from an archaebacterial ancestor (eg. pol 2 and pol 3) are marked by an alternating dot-dash line and a dashed line, respectively

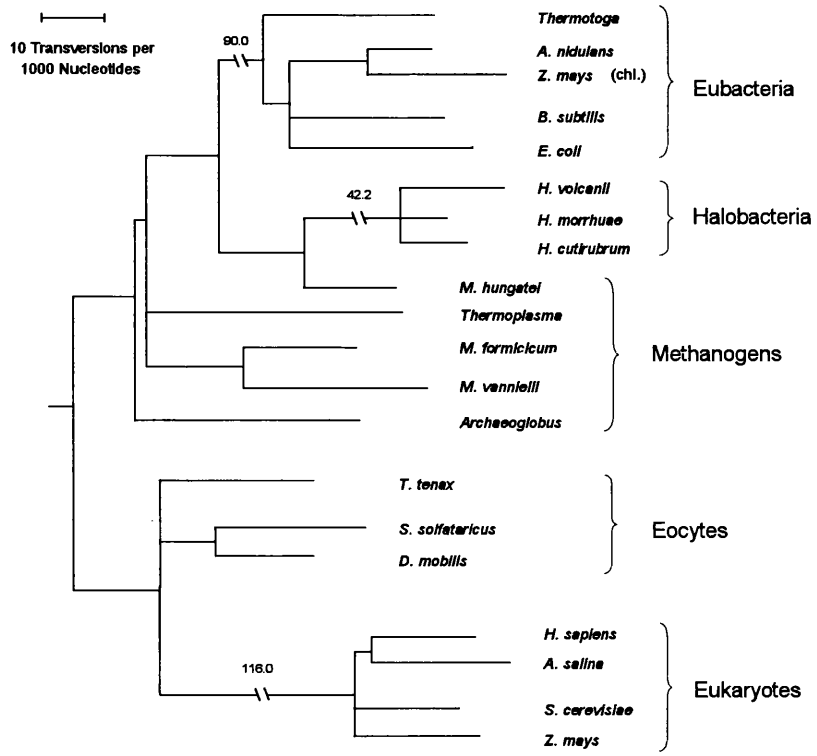


Fig. 1.5 (b) Phylogenetic tree, unrooted, proposed by Lake [1991] from 16S and 18S rRNA sequences, using an evolutionary parsimony algorithm.

1.2.2 ARCHAEAL PHENOTYPES

The Archaea encompass three basic phenotypes: methanogenic, halophilic and thermophilic, with some species possessing more than one phenotype, for example the thermophilic methanogens. These are grouped together as the Euryarchaeota (methanogens, extreme halophiles and thermophiles) and the Crenarchaeota (sulphur-dependent thermophiles). Woese proposes that the Euryarchaeota are evolutionarily closer to the Eucarya. Recent developments have led to the possibility of determining rRNA sequences from organisms that have not yet been cultured, using the technique of polymerase chain reaction (PCR) [Barns *et al.*, 1994]. The results have expanded the diversity of archaeal species and also suggested that the divide between the Euryarchaeota and the Crenarchaeota may be less well defined than originally thought. With the use of the PCR technique it

has also been revealed that the Archaea are abundant in particular marine habitats where the ambient temperatures range from -1.5°C to 15°C [Fuhrman *et al.*, 1992; DeLong *et al.*, 1994], thus forming a novel phenotypic archaeal group, the psychrophiles, in addition to those three phenotypes mentioned above.

The Methanogens:

The methanogens are the most common of all the Archaea, found in most anaerobic habitats such as soil, sewage sludge and ruminant guts. They are also the most diverse with mesophilic, thermophilic and halophilic representatives. All methanogens are characterised by their possession of a highly complex and novel methanogenesis pathway for the reduction of simple carbon compounds such as methanol, CO, CO_2 and acetate to methane [for reviews see Daniels, 1993; Schonheit, 1993; for reviews of archaeal metabolic pathways see Danson, 1988; Danson & Hough, 1992].

The Halophiles:

The halophiles require high sodium chloride concentrations for growth and are found in hypersaline environments, such as solar salterns, where the extracellular concentration of sodium salts is between 2-5 M. Some halophiles have an additional requirement for high pH and can be found in soda lakes (pH 9-10). The intracellular conditions are isotonic with the extracellular conditions, but are potassium chloride instead of sodium chloride, and consequently halophilic proteins have an obligate requirement for high ionic strength in order to function optimally.

The Thermophiles:

The thermophiles are a diverse group of organisms, with a wide range of physiological and biochemical properties [reviewed in Gottschal & Prins, 1991]. All thermophiles are characterised by their high growth temperatures. Moderate thermophiles have an optimum growth temperature of at least 50°C, extreme thermophiles of at least 65°C, and hyperthermophiles of at least 85°C. The aerobic

sulphur-dependent thermophiles can also be acidophilic and grow in acidic conditions as low as pH 1. The most common thermophilic habitats are between 40°C-70°C, such as solar-heated soil and decaying vegetation. However, most characterised archaeal thermophiles are hyperthermophiles and are usually found in geothermally heated springs and pools.

1.2.3 BIOTECHNOLOGY OF THE ARCHAEA

One of the major reasons for the increasing interest in the study of the Archaea is the opportunities that these organisms can contribute to the development of biotechnology [for reviews see Hough & Danson, 1989; Danson *et al.*, 1992; Cowan, 1992]. As a result of the nature of the extreme environments that the Archaea are capable of inhabiting these organisms possess a wealth of adapted phenotypic characteristics that can be exploited. Indeed, many Archaea have already been exploited over the past decade; for example, the use of anaerobic methanogens to produce the fuel, methane, is an established technology, as is the use of halophiles in food processes such as the production of fish paste, and many stable enzymes from the thermophiles are currently employed in high-temperature applications.

1.2.4 *THERMOPLASMA ACIDOPHILUM*

The Archaeon *Tp. acidophilum* is the source organism of the thermostable citrate synthase required to investigate protein stability.

Tp. acidophilum is a thermoacidophilic Archaeon, first isolated from a coal refuse pile in Indiana, USA [Darland *et al.*, 1970]. Since then it has been found in naturally-occurring hot springs and sulphataric fields [Segerer *et al.*, 1988]. The organism grows between 45-62°C, with an optimum growth temperature of 57°C, and between pH 1-2. From studies of 16S rRNA sequences it was proposed that

Tp. acidophilum groups with the Euryarchaeotes, even though the Archaeon phenotypically resembles those organisms of the Crenarchaeota group [Yang *et al.*, 1985]. More recent studies of 23S rRNA suggests that *Tp. acidophilum* actually stems from an ancient, and possibly unique, divergence within the archaeal domain (Fig. 1.6) [Ree *et al.*, 1993]. This result is consistent with an analysis of DNA-dependent RNA polymerases from *Tp. acidophilum* and other Archaea [Klenk *et al.*, 1992]. Clearly, the true phylogenetic position of *Tp. acidophilum* within the Archaea will not be elucidated without further comparative analyses.

Since *Tp. acidophilum* grows optimally at moderately-high temperatures and in acidic conditions, it must possess features that allow the organism to withstand these extremes. The organism does not have a rigid cell wall, and is rather amoeboid in nature; It may be that this characteristic is a nutritional adaptation, where flexibility of the cell membrane allows maximum contact and utilisation of insoluble substrates such as elemental sulphur [Searcy, 1986]. The lipids found in the cell membrane of *Tp. acidophilum* are typical of those in the Archaea, in that they are mainly isoprenyl glycerol ether-linked lipids, compared to the usual fatty acid glycerol ester-linked lipids of the Eucarya and the Bacteria [reviewed in Gambacorta *et al.*, 1994]. In addition, all glycerol ethers in the Archaea possess the unusual 2,3-*sn* stereochemistry. This lipid composition reflects the essential stabilising adaptation of the organisms to extremophily and also acts as a useful marker for archaeal taxonomy.

Although *Tp. acidophilum* grows in very acidic conditions, the internal pH of the cytoplasm is maintained at pH 6. The organism does not possess the proton-translocating ATPase, which is used in other prokaryotes and the Archaea to regulate the internal pH. It is thought that the absence of the enzyme could be due to the extreme nature of its habitat and instead the internal condition of the cell is maintained with a short membrane-bound electron transport chain [Searcy, 1986].

The DNA of *Tp. acidophilum* has a G+C base composition of 46 % [Christiansen *et al.*, 1975] and is involved in interactions with a histone-like protein, HTa, reducing

the susceptibility to thermal denaturation [Searcy, 1975; Stein & Searcy, 1978; Delange and Williams, 1981].

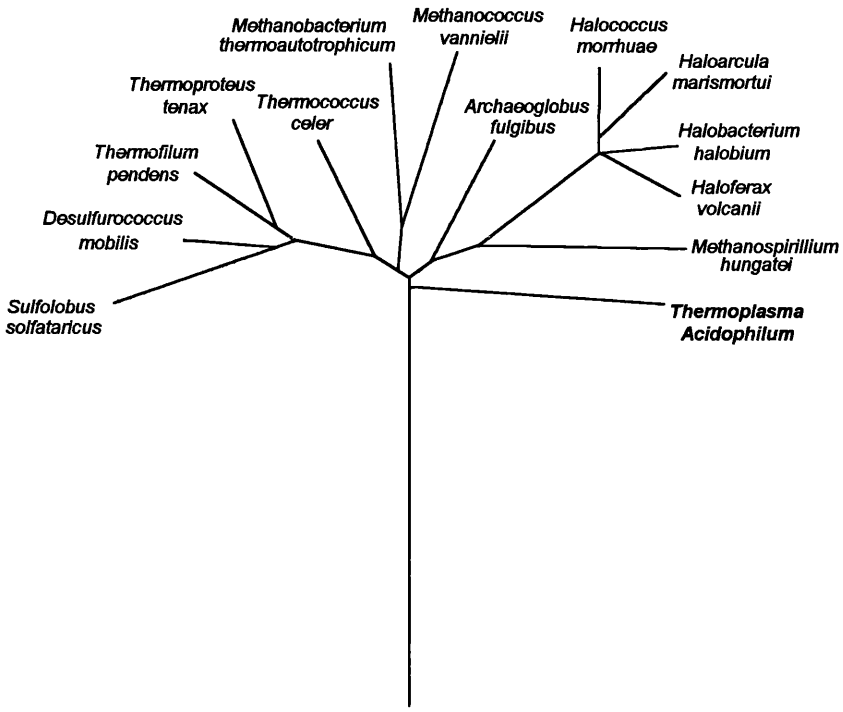


Fig. 1.6 Phylogenetic tree of the Archaea based on 23S rRNA sequences. The sequence of the 23S rRNA from *B. subtilis* was used as an outgroup to root the tree [Ree *et al.*, 1993].

The thermostabilising features of *Tp. acidophilum* proteins were not well documented, until recently. A limited number of proteins from this organism have been sequenced, for example, Hta [Delange and Williams, 1981], elongation factors-1 and -2 [Tesch & Klink, 1990; Pechmann *et al.*, 1991], ferredoxin [Wakabayashi *et al.*, 1983], glucose dehydrogenase [Bright *et al.*, 1993] and citrate synthase [Sutherland *et al.*, 1990]. Sequence comparisons of these proteins shed little light on the putative stabilising properties; however, recently, the crystal structures of glucose dehydrogenase and citrate synthase from *Tp. acidophilum* were determined and have been used in structural comparative analyses to identify possible thermostabilising features [John *et al.*, 1994; Russell *et al.*, 1994].

1.3 PROTEIN THERMOSTABILITY

Defining protein thermostability, and usually the accompanying stability to solvents and denaturants, has an importance to the development of biotechnological applications, and also to the fundamental understandings of the forces that play a part in protein folding and inactivation. The conformational stability of a protein is defined as the difference in free energy between the folded and unfolded states, with the native conformation possessing the global free energy minimum relative to all other states. The difference in the values of free energy is low, between 5-15 kcal/mol [Pace, 1990], surprisingly so since it has been possible to construct small proteins with higher conformational stability [Regan & Degrado, 1988]. This seemingly low stability for natural proteins may be connected with their *in vivo* turnover and regulation.

Protein conformational stability is determined by the relationship between opposing energetic forces represented by enthalpic and entropic terms. The enthalpic term includes the forces that play a major role in stabilising the native state of the protein and include van der Waals interactions, salt bridges, hydrophobic interactions, hydrogen bonds, and covalent interactions such as disulphide bonds. The entropic term includes conformational entropy which is a measure of the mobility within a molecule, and as such is the major force opposing protein folding [reviewed in Dill, 1990; Jaenicke, 1991].

Although the research into protein stability is progressing, we are still far away from defining a group of properties that are unique to thermostable proteins. However, some general trends, termed 'traffic' rules, are emerging. It appears that in the majority of cases so far investigated, stability is not the function of a single feature but instead an accumulation of properties that all make small contributions.

There are three main areas of research into protein stability: (a) Comparative analyses between mesophilic and thermophilic protein sequences have been conducted, the results of which form the basis of the majority of the traffic rules that have been proposed; (b) The 3D structures of mesophilic and thermostable

homologs also have been used in comparative analyses, verifying and adding to the traffic rules; (c) Finally, denaturation studies, on both wild-type proteins and site directed mutants, mainly of small monomeric structures, have been used to both identify new and test proposed features of the traffic rules. All methods are insufficient alone, but together they complement each other and can provide a fuller characterisation of stability in a given protein.

1.3.1 COMPARATIVE SEQUENCE ANALYSES

The most extensive comparative analysis carried out on amino acid sequences was conducted by Menendez-Arias and Argos [1989], involving 65 sequences from six different protein families, with at least one thermostable representative in each family and also the 3D structure known for at least one member. The results of the analysis describe the ten most frequent amino acid exchanges going from mesophile to thermophile, together with the change in hydrophobicity induced and the respective preferred location of the exchange (Table 1.4). In short, as thermostability increased so did the hydrophobicity of the protein, whilst the flexibility decreased. The most sensitive areas to these exchanges were α-helices, especially in exposed positions, and at domain/subunit interfaces.

There appear to be three prominent trends in the exchanges. Most markedly, the Ala content, particularly of helices, was increased. Ala has been shown to be the amino acid with the highest helical propensity and is known to stabilise these secondary structures [O'Neil & Degrado, 1990; Horovitz *et al.*, 1992]. The mechanism of helix stabilisation is as a result of the restriction in the conformational freedom of Ala in the unfolded state, such that the loss in entropy upon helix formation is smaller than for any other residue [O'Neil & Degrado, 1990]. A second trend is the exchange of Lys to Arg in exposed regions of the protein; this pattern is thought to be connected to the increased solvation of Arg compared to Lys, and controlled guanidination of solvent-exposed Lys residues has been demonstrated to increase the thermostability of some proteins [Qaw & Brewer, 1986]. Finally, there

are a number of exchanges that involve the addition of methyl groups to the residue side chain, for example Ser to Thr, and these are thought to be stabilise the folded conformation through increased hydrophobic interactions and also by decreasing the change in entropy upon folding through hydrogen bonding to the main chain in the unfolded state. For the same reasons, the exchange from Asp to Glu is thought to be stabilising.

AMINO ACID EXCHANGE	EFFECT ON HYDROPHOBICITY	PREFERRED LOCATION
Lys to Arg	+	helix
Ser to Ala	+	helix
Gly to Ala	+	helix & sheet
Ser to Thr	+	sheet & coil
Ile to Val	-	helix
Lys to Ala	+	helix & sheet
Thr to Ala	+	sheet
Lys to Glu	+	coil
Glu to Arg	+	helix
Asp to Arg	+	helix

Table 1.4 The ten most frequent amino acid exchanges found from comparative analyses of six families of mesophilic/thermophilic proteins. The preferred position for each exchange is indicated, as is the change in hydrophobicity (+ increased, - decreased) [Menendez-Arias & Argos , 1989].

Another statistical analysis resulted in the development of the aliphatic index, defined as the relative volume of a protein occupied by aliphatic side chains [Ikai, 1980]. It was found that the index is higher in thermophilic proteins in contrast to mesophilic homologs.

There are other features that do not appear in the top ten exchanges listed above but have been supported by experimental evidence from irreversible-denaturation studies of proteins. These involve the decrease in Asn and Gln, which are susceptible to deamidation at elevated temperatures [Stephenson & Clarke, 1989; Geiger & Clarke, 1987], in Asp residues, which can cause cleavage of the

polypeptide in weakly acidic conditions [Zale & Klibanov, 1986; Ahern & Klibanov, 1985], and the decrease, or total absence, of Cys residues, which are susceptible to oxidation at elevated temperatures [Zale & Klibanov, 1986].

The sequences used in the analyses above all originated from either eucaryal or bacterial organisms, with the highest optimum growth temperature of only 80°C. Since the majority of characterised thermophilic Archaea are actually hyperthermophiles with optimum growth temperatures in excess of 90°C, more archaeal sequences, in particular with hyperthermophilic representatives, are required.

The value of the information gleaned from sequence analyses, such as the one carried out by Menendez-Arias & Argos [1989], is only limited since it fails to take into account the role that a specific residue or entire secondary structure may play in the context of the 3D structure of the protein molecule. This information can be gathered only from comparisons of thermophilic and mesophilic 3D structures.

1.3.2 COMPARATIVE STRUCTURAL ANALYSES

There are a limited number of proteins for which the 3D structures have been determined for both mesophilic and thermophilic homologs. In all cases, comparative structural analyses have revealed that the topologies of the homologous proteins are very similar, and as such have proved useful in identifying putative thermostabilising features, pinpointing specific interactions between residues and assessing the stability of regions of the molecule. In all the models, the general trend from mesophile to thermophile involves a few additional interactions, these taking the form of salt bridges in all but one example, and also hydrophobic interactions in some of the thermophilic proteins.

From comparisons of the crystal structures of malate dehydrogenase (MDH) from pig and *Thermus thermophilus*, the thermostability of the latter protein is thought to be due to a few extra internal ionic interactions [Kelly *et al.*, 1993]. The bacterial

enzyme structure has four additional ion pairs, three that increase interactions at the subunit interface in the dimeric molecule and one at the domain interface. Other minor features that could play a role in the thermostability of MDH include α-helix stabilisation with an increase in the number of Ala residues and with a charged residue to cancel the macro-dipole across the secondary structure, and a general increase in the hydrophobicity accompanied by a decrease in the flexibility of α–helices and interdomain regions.

On the other hand, from the comparison of the crystal structures of ribonuclease H (RNaseH) from *E. coli* and *Thermus thermophilus*, the main thermostabilising feature is thought to be increased levels of aromatic-aromatic interactions [Ishikawa *et al.*, 1993(a)]. The latter enzyme possesses a cluster of eight aromatic residues, a feature which has been shown to occur in thermostable proteins [Burley & Petsko, 1985]. In addition, this protein does appear to have five more salt bridges, some of which stabilise the α-helices through cancelling the macro-dipole. A more significant change in RNAseH from mesophile to thermophile is the exchange of a left-handed Lys residue to a Gly, effecting the removal of steric hindrance caused by the Cβ atom of the Lys.

Neutral protease is the protein mentioned previously, in which additional salt bridges do not appear to play a role in increasing its thermal stability. Instead from the limited comparison of the crystal structures of the enzymes from *B. cereus* and *B. thermoproteolyticus*, it would seem that protein stability is the result of extra hydrogen bonds and an exchange of Gly to Ala residues within the helical structures [Pauptit *et al.*, 1988].

Many more stabilising features were suggested for the enzyme phosphoglycerate kinase (PGK), from a comparison between the crystal structures of yeast and *B. stearothermophilus* PGK [Davies *et al.*, 1993]. The most significant difference between the two protein structures is the number and position of ion pairs. The thermophilic PGK possesses an additional 15 ion pairs, ten of which were shown to stabilise the secondary structures within the N-terminal domain, and this is thought to be particularly important since this domain is the least stable part of the yeast

enzyme. There are amino acid exchanges from Ser and Thr to Ala, Glu and Asp, and also Lys to Arg, all of which are in agreement with the traffic rules proposed by Menendez-Arias & Argos [1989]. Finally, many of the loop regions joining the secondary structures are shorter in the thermostable PGK than in the mesophilic protein. With the addition of a 'closing' effect caused by a shift in domain orientations, the entire protein appears to be more compact and less flexible in the thermophile.

Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) from lobster, *B. stearothermophilus* and *Thermotoga maritima* have been crystallised and compared in order to identify features conferring both moderate and hyper-thermostability [Walker *et al.*, 1980; Korndorfer *et al.*, 1995]. The three organisms range from psychrophilic, growing at temperatures below 5°C, to hyperthermophilic, respectively. From the limited analysis of this protein model, it was suggested that thermostability is the result of an increased number of ion pairs at the surface of the molecule and increased hydrophobic interactions at the subunit interface of the four monomers. Hyperthermostability may be attributed to the exchange of a Val residue with unfavourable bond angles, to a more stable Gly residue, and also the exchange of Gly to Ala, Ser to Thr, and Ile to Val residues, all of which are in agreement with the traffic rules proposed by Menendez-Arias & Argos [1989]

One other protein model has been used in comparative structural analyses, that being citrate synthase, for which the crystal structures of pig and *Tp. acidophilum* have been compared [Russell *et al.*, 1994]. The results of this analysis will be discussed thoroughly in section 3.3.2, since they only became apparent 22 months into the work described in this thesis.

There are a few more thermostable proteins that have had their 3D structures solved, those being glucose dehydrogenase (GDH) from *Tp. acidophilum* [John *et al.*, 1994], isopropyl malate dehydrogenase (IPMDH) from *T. thermophilus* [Imada *et al.*, 1991], and aldehyde ferredoxin oxidoreductase (AFOR) from *P. furiosus* [Chan *et al.*, 1995]. For all these proteins there is no mesophilic structural equivalent and instead they have been used in two other types of comparative analyses: Either the

thermophilic structure has been compared to the structure of a mesophilic relative within the protein family, for example the crystal structure of IPMDH was compared to that of isocitrate dehydrogenase from *E. coli* [Imada *et al.*, 1991], or, the amino acid sequence of the thermophilic protein has been compared to sequences from mesophilic homologs, as is the case with AFOR [Chan *et al.*, 1995]. The protein rubredoxin is a slightly different case, where the secondary structure of the molecule from *P. furiosus* has been defined by NMR, and used in limited analyses with the 3D crystal structure of the protein from *Clostridium pasteurianum* [Blake *et al.*, 1991]. Again, the general trend appears to be an increase in ionic interactions, internal hydrophobic interactions, and secondary structure stabilisation through appropriate residue exchanges or additional bonding within the structure.

Other proteins have had the amino acid sequence determined for a thermophilic homolog and have had this aligned to the known mesophilic 3D structure of the same protein. Examples include the proteins ferredoxin [Perutz & Raidt, 1975], triosephosphate isomerase [Rentier-Delrue *et al.*, 1993] and glutamate dehydrogenase [Britton *et al.*, 1995], where the results of the comparison tend to be the same as described above but the emphasis is more on residue exchanges.

These latter comparative analyses are not as informative or as valid as the true structural comparisons. However, in the absence of 3D structures of both mesophilic and thermophilic homologs of the same protein, the comparison to the structures of related proteins, or sequence comparisons and homology modelling, are adequate for speculating which features may be important for conferring protein thermostability.

1.3.3 SITE-DIRECTED MUTAGENESIS

Site directed mutagenesis, together with denaturation and refolding studies, have enabled the testing of the traffic rules that were generated from the sequence and structural analyses of mesophilic and thermophilic proteins. In addition, other different features have been mutated and found to affect protein stability and so

have been added to the list of rules. The majority of the research has been carried out on small monomeric proteins, such as ribonuclease from *B. amyloliquefaciens* (barnase), bacteriophage T4 lysozyme, *E. coli* ribonuclease H1 and nuclease from *Staphylococcus*.

Extensive analyses on large numbers of mutants have suggested that the effects of most mutations are dependent upon the structural context. For example, Rennell *et al.*, [1991] created 2015 site-directed mutants of bacteriophage T4 lysozyme, mutating 163 of the 164 residues into 13 different amino acids in turn. The results showed that only 261 of the mutations had a detrimental effect upon the stability of active protein. Almost all of these mutations occurred in regions of the enzyme that had low mobility and/or low solvent accessibility. These results supported the work of Alber *et al.*, [1987], again on T4 lysozyme, who proposed that residues with high temperature factors in the crystal structure and usually surface exposed residues are relatively tolerant to mutations.

The majority of the features that have been identified as affecting protein stability involve changing the type and number of interactions, in both the folded and unfolded conformations. These mutational studies will therefore be described in terms of the types of interactions investigated.

1.3.3.1 IONIC INTERACTIONS

The main message from mutational analyses investigating ionic interactions is that engineered salt bridges on the surface of the protein make little contribution to the stability of the folded conformation. This result was found with the tetrameric GAPDH [Tomschy *et al.*, 1994], barnase [Sali *et al.*, 1991], and T4 lysozyme [Dao-Pin *et al.*, 1991]. The range of changes in ΔG encompassed 0.1-0.25 kcal/mol. Crystal structures of the T4 lysozyme mutants revealed that the engineered charged residues had such a high freedom of rotation that the ion pair did not come together to bond at all, and in fact it was suggested that the entropic cost of localising a pair of solvent-exposed charged groups on the surface of a protein largely offsets the

interaction energy expected from the formation of a defined salt bridge. The same theory applies to the disruption of exposed ionic interactions in barnase, where there is little difference between the stability of the mutant and the wild-type protein [Serrano *et al.*, 1990]. In contrast, the disruption of a buried salt bridge in T4 lysozyme was found to destabilise the protein, and was calculated to contribute 3-5 kcal/mol to the conformational stability of the folded state [Anderson *et al.*, 1990].

1.3.3.2 HYDROGEN BONDS

The contribution of hydrogen bonds to the native states of proteins has been investigated by disruption of the bonds, either mutating one of the residues involved in a pair or both. On the basis of urea and thermal unfolding studies of 12 mutants of ribonuclease T1 (RNase T1) with disrupted hydrogen bonds, Shirley *et al.* [1992] were led to conclude that the hydrogen bonds make a significant contribution to the conformational stability of the protein, so much so that they parallel that made by hydrophobic interactions. They found the average decrease in stability to be equal to 1.3 kcal/mol per hydrogen bond. To complement this work, studies of the hydrogen bonding patterns between the enzyme tyrosyl-tRNA synthase and its substrate have demonstrated that hydrogen bonds involving charged residues have stronger binding energies (3.5-4.5 kcal/mol) than those involving uncharged amino acids (only 0.5-1.5 kcal/mol) [Fersht *et al.*, 1985].

1.3.3.3 DISULPHIDE BONDS

Covalent interactions through disulphide bonds can have dramatic effects upon the stability of proteins. The most remarkable study investigating the effect of disulphide bond creation involved the protein thymidylate synthase from the organism *Lactobacillus casei*, where two symmetrically-related disulphide bonds were engineered at the subunit interface of the dimer [Gokhale *et al.*, 1994]. The mutant had a huge increase in stability; whereas the wild-type protein precipitates at 52°C, the mutant is still soluble and retains secondary structure even at 90°C. This

is further illustrated by the appreciable level of activity that the mutant thymidylate synthase displays at 65°C. The effect of creating extra covalent interactions in the enzyme T4 lysozyme was not quite so stark as that in thymidylate synthase. Four individual mutants were engineered to contain a disulphide bond, with the increase in thermal stability ranging from 5-10°C [Matsumara *et al.*, 1989]. The reason for the relatively small increase in stability is probably due to additional strain within the molecule increasing the change in conformational entropy from the unfolded to native states.

1.3.3.4 HYDROPHOBIC INTERACTIONS

Since hydrophobic interactions are thought to be the main driving force of protein folding, many investigators have tried to quantify the effect of these interactions upon the conformational stability of the folded state. Most studies have taken the form of investigating cavity-creating and cavity-filling mutants. From the urea and thermal unfolding studies of 72 aliphatic side chain mutants in four proteins, Pace [1992] has found the average change in ΔG to be 1.3 kcal/mol per methyl group, and this is supported by mutational work on barnase [Serrano *et al.*, 1992(a); Kellis *et al.*, 1988]. However, in the case of some cavity-creating mutants, the change was found to be larger than the 1.3 kcal/mol suggested, and instead it was proposed that the discrepancy in the destabilisation was determined by the ability of the residues lining the cavity to collapse and minimise the size of the hole [Eriksson & Matthews, 1992; Shortle *et al.*, 1990].

In addition, Ishikawa *et al.* [1993(b)] found that, with cavity-filling mutants of RNase HI, the change in stability did not increase to the degree that was predicted from those studies just described, apparently due to non-maximal van der Waals contacts within the cavity. It was also discovered that Leu was preferred to Ile to fill cavities, since Leu is able to adopt a wider range of dihedral angles and thus maximise potential interactions. In contrast, the cavity-filling mutants of T4 lysozyme created by Karpusas *et al.* [1989] were actually destabilised, due to strained conformations of the mutated residues.

Increasing hydrophobic interactions at the subunit interface in multimeric proteins has been shown to lead to an increase in the thermal stability of the molecule [Kotik & Zuber, 1993; Kirino *et al.*, 1994; Kallwass *et al.*, 1992]. In the case of the already thermostable lactate dehydrogenase from *B. stearothermophilus*, an Arg residue within a helix at the interface was mutated to Tyr, an amino acid usually thought of as a helix-breaker, and the melting temperature of resultant protein was 9.4°C higher than for the wild-type enzyme [Kallwass *et al.*, 1992].

1.3.3.5 EFFECTS ON THE CONFORMATIONAL ENTROPY

The effect of changes in the conformational entropy of the unfolded state of a protein upon the stability of the folded state has been investigated. Two mutants of bacteriophage T4 lysozyme were engineered, changing Gly 77 to Ala and Lys 60 to Pro; both were accompanied by a stabilising decrease in ΔG (of 0.4 and 0.8 kcal/mol, respectively) [Matthews *et al.*, 1987]. The small rise in protein stability was thought to be the result of a decrease in the change in conformational entropy upon folding due to a reduced degree of freedom of the amino acids Ala and Pro in the unfolded state, compared with the wild-type residues. Ishikawa *et al.* [1993(c)] conducted a similar study on RNase HI, where a His residue situated in a loop region was changed to a Pro, without affecting the native backbone structure. The result was a 4.1°C increase in the melting temperature of the protein.

In contrast, Bowler *et al.*, [1993] destabilised the native state of iso-1-cytochrome by increasing the stability of the unfolded conformation; this was achieved through additional hydrophobic interactions that demonstrated that the unfolded state was not a true random coil structure and that the hydrophobic residues were partially shielded from full solvation. The fact that the hydrophobicity of the mutated residues correlated with the change in ΔG supported the theory concerning the nature of the unfolded conformation.

1.3.3.6 HELIX STABILISATION

The factors that affect helix stability, and in turn the stability of the whole molecule, have been studied thoroughly. It appears that helical content is important although this depends on the position of the residue within the helix and the position and exposure of the helix within the molecule. Ala has been shown to be the amino acid with the highest helical propensity, due to its hydrophobicity and appropriate dihedral angles [O'Neil & Degrado, 1990; Horovitz *et al.*, 1992], although the stabilising effect is context-dependent and site-specific [Serrano *et al.*, 1992 (b) (c)]. Residues that have been shown to destabilise helical structures include Gly at internal positions, due to the greater freedom of bond rotation, Pro, due to the inability of the amino acid to adopt the optimum main chain angles for an α-helix and to the disruption of the intrahelical hydrogen bonding network, and also maybe residues with branched side chains, because the loss of entropy is too great upon helix formation due to the restricted number of side chain conformations [Horovitz *et al.*, 1992]. On the other hand, in solvent exposed helices, Gly residues are known to have a stabilising effect if they are positioned at the termini, in particular the C-terminus, since they allow the exposure of the amino and carboxy groups available for solvation [Horovitz *et al.*, 1992; Serrano *et al.*, 1992(c)]. Amino acids that are capable of bonding to the main chain in the unfolded state, for example Ser, Asp and Tyr, also have been found to be destabilising since the change in entropy upon folding is increased [Horovitz *et al.*, 1992].

There are a variety of other amino acids that are capable of stabilising helices if they are found at the termini; these are known as capping residues and include Asp, Asn, Ser and Thr at the N-terminus of the helix, and His, Arg, Lys and Asn at the C-terminus [Hol, 1985; Serrano *et al.*, 1992(b)]. The mechanism of capping at the N-terminus involves hydrogen bonds between the side chain of the N-cap (residue N+0) to the main chain amino group of N+3. In addition, Asp is particularly favourable since it can interact with the macro-dipole across the helix. The mechanism of capping at the C-terminus is similar, with most of the C-cap residues possessing positive torsion angles that allow the side chain of the residue to interact

with main chain carboxy groups at the end of the helix; also basic amino acids can interact with the helix dipole. Close scrutiny of conflicting data, from mutational studies in T4 lysozyme [Bell et al., 1992], has also revealed that it is possible for surrogate caps to stabilise a helical structure [Serrano *et al.*, 1992b].

1.3.4 AN OVERVIEW OF PROTEIN THERMOSTABILITY

From all the investigations into protein thermostability it is apparent that the factors responsible are context specific and, with a few exceptions, it appears that conformational stability is the accumulation of small improvements in intermolecular interactions. Overall, the thermostable protein molecule appears to be more rigid and less flexible at mesophilic temperatures, although the protein attains a flexibility at its operating temperatures, similar to that of the mesophilic enzyme at its operating temperature; this is almost a prerequisite since with enzymes, activity depends on conformational and amino acid side-chain mobility.

All the thermostabilising factors described above are of an intrinsic nature. However, there are some proteins, especially those from hyperthermophiles, that are not as stable *in vitro* as the growth temperature of the source organism would predict. It is thought that the high concentrations of certain intracellular molecules, such as particular polyalcohols, polyamines and ions, found in these organisms may provide an extrinsic mechanism of stabilisation [De Cordt *et al.*, 1994; Zellner and Kneifel, 1993; Hensel and Jakob, 1994; Scholz *et al.*, 1992].

At the onset of this work, virtually all of the studies into protein thermostability were conducted on proteins from thermophilic Bacteria. Through analysing proteins, such as citrate synthase, from the thermophilic Archaea, such as *Tp. acidophilum*, it should become apparent if the Archaea use similar or novel means of achieving the same goal of protein stability.

1.4 AIMS OF THE PROJECT

The ultimate aim of this project was to characterise protein thermostability in the enzyme citrate synthase from *Tp. acidophilum*. Citrate synthase was been chosen as the model with which to study thermostability for a number of reasons :

(1) Citrate synthase is a ubiquitous enzyme found in almost every known organism. Therefore, there are representatives of the enzyme from a whole range of different biotopes, such as from mesophilic conditions to extremes of temperature and salinity. There are also representatives found in the three domains of life, enabling the characterisation of protein thermostability in the Eucaryotes, the Bacteria and the Archaea. From a comparison of these three systems, it should become apparent as to whether phylogenetically-distinct organisms have employed similar means for achieving the common goal of protein stability.

(2) Citrate synthase is a dimeric protein where both subunits are essential for catalytic activity. Therefore, integrity of the dimeric nature of the molecule at the elevated temperature at which *Tp. acidophilum* grows, can be determined from a study of the interactions at the subunit interface. The information from this investigation will add to the limited data that are available for multimeric protein structures.

(3) There is an expanding database of information concerning citrate synthase to facilitate the comparative analyses.

The work that is presented here is divided into two main areas: the identification of features that could play a role in conferring protein thermostability and the testing of those theories by site directed mutagenesis. Chapter 3 describes the homology modelling of a 3D structure of citrate synthase from *Tp. acidophilum*, together with comparative analyses involving the amino acid sequences and 3D structures of the protein. Features are identified that may play a role in the thermostability of the protein. A few of these features have been investigated by site-directed mutagenesis in Chapters 4 to 6.

CHAPTER 2

MATERIALS AND METHODS

All the chemicals used were of analytical grade or of the finest grade commercially available. Unless otherwise stated, chemicals and reagents were obtained from BDH Chemicals, Poole, UK; Sigma Chemicals, Poole, UK; Fisons, Loughborough, UK and Fluka Chemicals Ltd, Gillingham, UK.

2.1 MOLECULAR MODELLING

All programs were run on the MicroVAX 3300 system, Silicon Graphics Indigo, or Evans and Sutherland ESV-10 work station.

Homology modelling and manipulations of a 3D structure of citrate synthase were achieved using the program O [Jones *et al.*, 1991]. The *mutate* functions within O enable the replacement, deletion and insertion of residues in the protein sequence. The *proleg* functions allow the structural manipulation of those residues. *Proleg* contains a database of information that is probed to reveal 20 structures of a specified number of residues between two given points, the most suitable of which is chosen.

The modelled structure was then energy minimised using the crystallographic package XPLOR Version 2.1 [Brunger, 1990]. This cyclic process involves the calculation of the total energy of the system as defined by a range of energy terms based on covalent bonds, bond angles, dihedral angles and non-bonded interactions. The conformation of the system is adjusted and the total energy recalculated until a minimum value is achieved.

The stereochemical quality of the protein structure was assessed using a number of programs. PROCHECK Version 2.1 [Laskowski *et al.*, 1993] generates a Ramachandran plot and highlights residues that have unfavourable stereochemistry with respect to side chain angles and main chain parameters. DSSP [Kabsch & Sander, 1983] defines the secondary structure and solvent accessibility of each residue within the 3D structure. The output file from DSSP is then analysed by the program ENVIRONMENTS [Luthy *et al.*, 1992] which highlights residues within unsuitable environments. Based on the information from the above programs, adjustments were made to the 3D structure and the energy minimisation repeated, until the unfavourable stereochemistry was reduced to a minimum.

2.2 MOLECULAR BIOLOGY

Hybaid Recovery™ Plasmid Midiprep kit was obtained from Hybaid Ltd, Middlesex, UK. Restriction endonucleases were supplied by New England Biolabs, Beverley, USA and Gibco BRL Life Sciences, Uxbridge, UK, as was 1kb DNA ladder. GENECLEAN® II kits were obtained from Bio 101 Inc., La Jolla, USA. Klenow fragment, T4 DNA polymerase, *Taq* DNA polymerase, T4 DNA ligase and T4 polynucleotide kinase were all supplied by Boehringer-Mannheim, Mannheim, Germany. Vent® DNA polymerase was obtained from New England Biolabs. Deoxynucleoside triphosphates (dNTP) were obtained from Pharmacia, Milton Keynes, UK. Altered Sites™ *in vitro* Mutagenesis System was supplied by Promega Corporation, Madison, USA. Sequenase® Version 2.0 DNA Sequencing kit was from United States Biochemical Corporation, Ohio, USA. Sequagel™ reagents were obtained from National Diagnostics, Atlanta, USA. $\alpha[^{35}S]$-ATP was obtained from ICN Biomedicals Ltd., High Wycombe, UK. *Hind*III-digested λ DNA, isopropyl-β-D-thiogalactoside (IPTG) and 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal) were obtained from Northumbria Biologicals Ltd., Cramlington, UK. Autoradiography film was supplied by the Fuji Film Company, Japan, and 3MM

Chromatography paper by Whatman Scientific Ltd., Maidstone, UK. Constituents of *E. coli* culture media were supplied by Difco Laboratories, Michigan, USA.

The vectors used in this study included pKK223-3, obtained from Pharmacia, pMEX8 and pJLA602, both supplied by MEDAC, Hamburg, Germany, and M13mp19 obtained from Stratagene Ltd., Cambridge, UK

The *E. coli* strains used were DH5αF' [F' σ80d*lac*ZΔM15 Δ(*lac*2YA-*arg*F)U169 *end*A1 *rec*A1 *hsd*R17($r_K^-m_K^+$) *deo*R *thi*-1 *sup*E44 λ^-*gyr*A96 *rel*A1], obtained from Gibco BRL Life Sciences, JM105 [*end*A *sbc*B15 *hsd*R4 *rps*L *thi* Δ(*lac-pro*AB) F'(*tra*D36 *pro*AB^+ *lac*1^q *lac*ZΔM15)], supplied by Pharmacia and MOB154 [*glt*A6 *gal*K30 *pyr*D36 *rel*A1? *rps*L129 *thi*-1 *sup*E44 *hsd*R4 *rec*A], a citrate synthase-deficient strain, kindly donated by D.O. Wood (University of South Alabama, USA).

2.2.1 GROWTH AND MAINTENANCE OF BACTERIAL STRAINS

The *E. coli* strains DH5αF' and JM105 were grown in LB liquid medium (1.0% (w/v) Bacto tryptone, 0.5% (w/v) yeast extract, 1.0% (w/v) NaCl) or on LB solid medium (includes 1.5% (w/v) Bacto agar), at 37°C unless otherwise stated. The strain was maintained on solid minimal medium [described by Sambrook *et al.*, 1989] supplemented with thiamin.HCl (0.2 μg/ml).

Unless otherwise stated, the citrate synthase-deficient *E. coli* strain MOB154 was grown at 37°C in LB liquid medium and on LB solid medium, both supplemented with 11 mM glucose. For selection purposes and maintenance the strain was grown at 37°C on solid minimal medium supplemented with thiamin.HCl (5 μg/ml), uracil (35 μg/ml), 11 mM glucose and 2 mM glutamate salt. To detect a MOB154 strain expressing an active plasmid-encoded citrate synthase gene, the bacteria were grown on the supplemented solid minimal medium but without the glutamate.

For the growth of *E. coli* strains expressing plasmid-encoded ampicillin or tetracycline resistance genes, the medium was supplemented with ampicillin (100 μg/ml) or tetracycline (15 μg/ml), respectively.

Long term storage of bacterial strains was in a 24% (v/v) glycerol solution at -70°C.

2.2.2 PREPARATION OF PLASMID DNA

2.2.2.1 SMALL SCALE - 'MINIPREP'

A volume, 1.5 ml, from an overnight culture of *E. coli* carrying the plasmid of interest was spun at 13,000 x *g* for 10 min. The cells were resuspended in 100 μl of solution consisting of 25 mM Tris.HCl, pH 8.0, 10 mM ethylenediaminetetra-acetic acid (EDTA), 50 mM glucose, and then lysed with 200 μl 0.2 M NaOH, 1% (w/v) sodium dodecyl sulphate (SDS). The alkali was neutralised with 150 μl 3 M potassium acetate, pH 4.8, on ice for 3-5 min and the SDS pelleted by centrifugation at 13,000 x *g* for 10 min. Plasmid DNA and RNA are precipitated by the addition of an equal volume of propan-2-ol followed by centrifugation at 13,000 x *g* for 15 min. The pellet of nucleic acids was then washed with 70% (v/v) ethanol, dried under vacuum and resuspended in 25 μl MilliQ water or TE8 buffer (10 mM Tris.HCl, 1 mM EDTA, pH 8.0).

2.2.2.2 MEDIUM SCALE - 'MIDIPREP'

The larger scale preparation of plasmid DNA was achieved using the Hybaid Recovery™ Plasmid Midiprep kit. The technique utilises the affinity of a silica-based matrix for DNA, and can purify up to 200 μg plasmid DNA from a 30 ml volume of *E. coli* culture carrying a high copy number plasmid.

2.2.3 ETHANOL PRECIPITATION OF DNA

DNA was precipitated out of solution by the addition of 2 volumes of ice-cold absolute ethanol, 0.1 volumes 3 M sodium acetate pH 5.2, followed by incubation at -20°C for at least 1 h. The DNA was then pelleted by spinning at 13,000 x *g* for 10 min, after which it was washed with 70% (v/v) ethanol and the spin repeated. The pellet was dried under vacuum and then resuspended in MQW or 2TE8.

2.2.4 DIGESTION OF DNA WITH RESTRICTION ENDONUCLEASES

Digestion of DNA with restriction endonucleases was carried out in the accompanying buffers and following the supplier's suggested conditions for optimum enzymic activity. If necessary, the samples were incubated with DNAse-free RNAse (0.1 μg/μl) to remove any RNA.

2.2.5 AGAROSE GEL ELECTROPHORESIS

DNA samples were analysed by agarose gel electrophoresis, separating linearised DNA species relative to their molecular sizes [described by Sambrook *et al.*, 1989]. The samples, in buffer (0.04% (w/v) bromophenol blue, 0.04% (w/v) xylene cyanol, 7% (w/v) sucrose), were loaded onto agarose gels (0.7% (w/v) agarose, 0.005% (v/v) ethidium bromide, made up in TAE (40 mM Tris.acetate, 1 mM EDTA)). TAE was used for tank buffer and the samples were electrophoretically separated using a constant voltage of 60-90 V for approximately 1 h. The size markers used were either *Hind*III-digested DNA or 1 kb Ladder. The gel was then visualised with UV trans-illumination.

2.2.6 PURIFICATION OF DNA FROM AGAROSE GELS

Samples of interest were loaded onto a 0.7% (w/v) LMP (low melting point) agarose gel in TAE buffer and electrophoresed at 30-40 V for 2-3 h. DNA was visualised using a low-power UV lamp and the relevant band was sliced out of the gel. The DNA was then purified from the agarose using the GENECLEAN® II kit following the manufacturer's instructions. The technique utilises a specially-formulated silica matrix that binds DNA.

2.2.7 END-FILLING COHESIVE ENDS OF DNA

A DNA sample, 200 μg, was mixed in a 30 μl volume with 3 μl T4 DNA ligase buffer, 2 U Klenow fragment, 0.05 U T4 DNA polymerase, 5 U T4 polynucleotide kinase and 12.5 mM each dNTP and incubated at 37°C for 20 min. The DNA was extracted twice using a 20 μl volume of phenol/chloroform/isoamyl alcohol (25:24:1 (v/v/v)) and then precipitated with ethanol as described 2.2.3.

2.2.8 LIGATION OF DNA

Vector DNA and insert DNA, approximately 0.1 μg of each, were ligated together using 4-8 U DNA ligase enzyme. Incubations were carried out in the ligase buffer supplied, at room temperature for 1 h followed by 37°C for 1 h. Either the whole or half of the ligation sample was then used in the transformation of competent *E. coli* cells.

2.2.9 PREPARATION AND TRANSFORMATION OF COMPETENT *E. COLI*

E. coli cells were made competent as described in the Altered Sites™ *in vitro* Mutagenesis System booklet, based on the method of Hanahan [1985]. Details of the transformation of competent *E. coli* also are described in the text.

2.2.10 TRANSFECTION OF *E. COLI* WITH M13MP19

A 0.01 volume of an overnight culture of *E. coli* strain DH5αF', together with a 0.002 volume of M13mp19 phage supernatant, were added to soft agar (LB medium, 7.5% (w/v) Bacto agar), melted and held between 45-50°C. The agar was then supplemented with 0.5 mM isopropyl-thiogalactoside (IPTG) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) (40 μg/ml) and poured onto LB solid medium where it was allowed to set. The plates were incubated at 37°C for 19 h after which time plaques of phage-infected *E. coli* could be seen on the lawn.

To obtain M13mp19 phage particles, a plaque from the lawn described above was used to inoculate 5 ml 2YT media (1.6% (w/v) Bacto tryptone, 1.0% (w/v) yeast extract, 0.5% (w/v) NaCl) supplemented with 0.5% (w/v) sucrose, together with 5 μl of an overnight culture of *E. coli* strain DH5αF'. The culture was incubated at 37°C for 4-5 h with vigorous shaking. Phage were released into the medium, and a supernatant was obtained by pelleting the *E. coli* cells at 13,000 x *g* for 10 min twice. The phage supernatant was stored at 4°C.

The replicative form of M13mp19 can be isolated following the mini-prep method described in 2.2.2.1. This form of DNA is a double-stranded circle and can be manipulated in the same manner as a plasmid.

2.2.11 ISOLATION OF SINGLE-STRANDED DNA

Isolation of single-stranded DNA was carried out on a large scale using 100 ml of infected culture, prepared as described in 2.2.10. The bacterial cells were pelleted and removed by centrifugation at 5000 x *g*, 4°C, for 20 min and then again for 5 min. The phage particles were precipitated from the supernatant by the addition of 0.2 volumes of 30% (w/v) polyethylene glycol (PEG) 8000, 1.6 M NaCl, and incubation at 4°C for 1 h. The precipitation was followed by centrifugation at 5000 x *g*, at 4°C, for 20 min. The viral pellet was resuspended in 500 µl TE8 buffer and any remaining bacterial cells were removed by centrifugation at 13,000 x *g* for 5 min. The phage were precipitated again with 200 µl PEG/NaCl solution, incubated at 4°C for at least 1 h, and were then collected by centrifugation at 13,000 x *g* for 10 min. The pellet was resuspended in 500 µl TE8, after which the protein coat of the phage particles was removed by extraction with 200 µl 25:24:1 (v:v:v) phenol/chloroform/isoamyl alcohol by vortexing for 1 min, standing on ice for 1h, vortexing again for 1 min and then centrifuging at 13,000 x *g* for 10 min. The top aqueous layer, containing phage DNA, was retained and the whole step was repeated, using shorter standing times, until a white precipitate no longer appeared at the interface between the two layers. Traces of phenol were then removed from the aqueous layer with 200 µl 24:1 (v:v) chloroform/isoamyl alcohol and the DNA was ethanol precipitated as described 2.2.3.

2.2.12 OLIGONUCLEOTIDE SYNTHESIS AND PREPARATION

Oligonucleotides were either custom made by Pharmacia or Severn Biotech Ltd, Kidderminster, UK, or were synthesised in-house on an Applied Biosystems 381A DNA Synthesiser (A.J. Wolstenholme, University of Bath).

Those oligonucleotides that were made in house were removed from the column matrix in 1 ml concentrated ammonium hydroxide at 55°C overnight. The beads were then

pelleted at 6000 x *g* for 2 min and removed. The sample was dried under vacuum for 3-4 h, after which it was resuspended in 300 μl MQW. The DNA was ethanol precipitated, as described 2.2.3, and resuspended in 1 ml TE8. The purity and concentration of the oligonucleotide was determined by measuring the absorbance at 260 nm and 280 nm [Sambrook *et al.*, 1989]; A ratio of A_{260}/A_{280} above 1.75 indicates that the sample is pure, and an absorbance of 1 unit at 260 nm equates to 50 μg/ml double stranded DNA or 37 μg/ml single stranded DNA.

2.2.13 DNA SEQUENCING

Single and double stranded DNA was sequenced using the Sequenase® Version 2.0 kit following the manufacturer's instructions. The sequencing protocol utilises Sequenase® Version 2.0 T7 DNA Polymerase and incorporates the chain termination method of Sanger *et al.* [1977]. The radio-isotope used was α[^{35}S]-ATP.

Prior to sequencing, the double stranded DNA was denatured. Denaturation was achieved by adding 0.1 volume of 2 M NaOH, 2 mM EDTA, and incubating at 37°C for 30 min. The solution was neutralised by the addition of 0.1 volume of 3 M sodium acetate pH 5.2, and then the DNA was ethanol precipitated as described 2.2.3.

The sequencing reactions were loaded onto a preheated 6% (w/v) acrylamide/bis-acrylamide gel, prepared from Sequagel™ reagents following the manufacturer's instructions. The electrophoretic separation was carried out in TBE buffer (90 mM Tris.borate, 2 mM EDTA) at a constant power of 52 W for 1-5 h. The gel was fixed in 10% (v/v) glacial acetic acid, 10% (v/v) methanol, for 20 min, dried onto 3MM Whatman Chromatography Paper, and then exposed to autoradiography film for 1-3 days, before developing using an Amersham Hyperprocessor.

2.2.14 SITE-DIRECTED MUTAGENESIS

The Altered Sites™ *in vitro* Mutagenesis System was used to create site-directed mutants of citrate synthase. This system utilises a phagemid, pALTER-1™ that produces single-stranded DNA upon infection with helper phage. The phagemid contains a multiple cloning site inserted into the β-galactosidase gene so that recombinant vectors can be detected by blue/white selection when plated on solid medium containing IPTG and X-Gal. There are genes for both ampicillin and tetracycline resistance on the plasmid, although the gene product for ampicillin resistance is inactive due to a frameshift. This feature forms the basis of the mutagenesis system; propagation of the plasmid and its recombinants is performed under tetracycline resistance but the mutagenesis procedure reinstates ampicillin resistance with a 'repair oligonucleotide' so that there is a 80-90% mutagenic efficiency when grown in the presence of ampicillin.

Site-directed mutagenesis, and all genetic manipulations therein, were carried out following the instructions supplied with the system.

2.3 PROTEIN BIOCHEMISTRY

Protogel™ was supplied by National Diagnostics Ltd. Low molecular weight markers for protein electrophoresis were obtained from Bio-Rad Laboratories, Herts., UK. Matrex™ gel Red A and ultrafiltration membranes (cutoff 10 kDa) were supplied by Amicon Ltd., Stonehouse, UK. Sephadex G-25 matrix was obtained from Sigma and coenzyme A from Calbiochem-Novobiochem, Nottingham, UK.

2.3.1 PREPARATION OF CELL EXTRACTS

Overnight cultures of *E. coli* were centrifuged for 10 min to pellet the bacteria; this was carried out at 6000 x *g* for volumes above 1.5 ml, and 13,000 x *g* for 1.5 ml volumes. The cells were responded in 2TE8 (20 mM Tris.HCl, 2 mM EDTA, pH 8.0) at a density of 0.01-0.2 g cells / ml 2TE8. The bacteria were lysed by sonication at 140 W on ice, using either one or five 10 s bursts for volumes below or above 1 ml, respectively. The cell debris was removed by centrifugation at 13,000 x *g* for 10 min.

2.3.2 ESTIMATION OF PROTEIN CONCENTRATION

Protein concentration was determined using the method of Bradford [1976]. A 0.9 ml volume of filtered Bradford reagent (0.01% (w/v) Coomassie G-250, 4.75% (v/v) ethanol, 8.5% (w/v) H_3PO_4 was mixed with 0.1 ml sample and incubated at room temperature for 10 min. The absorbance at 595 nm was measured and the protein concentration calculated from a calibration curve of bovine serum albumin (BSA) standard solutions ranging from 0-25 μg/ml.

2.3.3 SODIUM DODECYL SULPHATE POLYACRYLAMIDE GEL ELECTROPHORESIS (SDS-PAGE)

Protein samples were analysed by SDS-PAGE based on the method of Laemmli [1970]. The resolving gel consisted of 10% (w/v) acrylamide (Protogel™), 0.375 M Tris.HCl pH 8.8, 0.1% (w/v) SDS, 0.025% (w/v) ammonium persulphate and 0.1% (v/v) N,N,N',N'-tetramethylethylenediamine (TEMED). The stacking gel consisted of 5% (w/v) acrylamide (Protogel™), 0.13 M Tris.HCl pH 6.5, 0.1% (w/v) SDS, 0.03% (w/v) ammonium persulphate and 0.001% (v/v) TEMED. To prepare the protein samples for electrophoresis an equal volume of loading buffer (0.18 M Tris.HCl pH 6.5, 2% (w/v)

SDS, 10% (w/v) glycerol, 2.6% (w/v) β-mercapto-ethanol, 0.005% (w/v) bromophenol blue) was added to the samples followed by an incubation at 100°C for 2 min. Low molecular weight markers were also run on the gel. Electrophoresis was carried out in tank buffer (0.025 M Tris.HCl pH 8.8, 0.096 M glycine, 0.1 % (w/v) SDS) at a constant current of 10 mA through the stacking gel and 20 mA through the resolving gel until the bromophenol blue dye front was at the lower edge of the gel. Following this, gels were stained for 30 min with Coomassie stain (0.5% (w/v) Coomassie Brilliant Blue R, 45% (v/v) methanol, 45% (v/v) acetic acid) and destained in 10% (v/v) methanol, 10% (v/v) acetic acid.

2.3.4 CITRATE SYNTHASE ASSAY

Citrate synthase activity was measured spectrophotometrically by the method of Srere *et al.* [1963]. The thiol group of the coenzyme A produced reacts with 5,5'-dithio-bis(2-nitrobenzoic acid) (DTNB) to release thionitrobenzoate which absorbs at 412nm and has an absorbance coefficient of 13,600 $M^{-1}cm^{-1}$. The standard reaction mixture contained 0.2 mM oxaloacetic acid, 0.15 mM acetyl coenzyme A, 0.2 mM DTNB and the enzyme sample, made up to a 1 ml volume with 2TE8 reaction buffer and incubated at 37°C.

Acetyl CoA was prepared as a 7.5 mM stock solution. To a solution of coenzyme A (10 mg/ml) on ice, a 0.2 volume of 1 M $KHCO_3$ was added followed by a 0.1 volume of fresh 1 M acetic anhydride. The solution was incubated on ice for 10 min to allow acetylation of the coenzyme A.

Enzyme assays were carried out in a Perkin Elmer Lambda 3B spectrophotometer and initial rates were calculated using the PECSS software supplied with the instrument. Specific activities of the samples were calculated as Units/mg protein, where 1U is defined as 1 μmol product formed/min.

2.3.5 PURIFICATION OF CITRATE SYNTHASE

Recombinant pig citrate synthase was purified from an *E. coli* MOB154 strain which was carrying the citrate synthase gene on the expression plasmid pKK223-3. A cell extract was prepared from a 2 l culture, as described in section 2.3.1. The cell extract was loaded onto a 20 ml affinity column of Matrex™ gel Red A, pre-equilibrated with 2TE8, at a rate of 1 ml/min. The column was washed with 2TE8 buffer, at 1 ml/min, until the absorbance at 280 nm reached a base line, after which citrate synthase was eluted with 50 ml elution buffer (1 mM oxaloacetate, 0.2 mM coenzyme A, 20 mM Tris.HCl, 2 mM EDTA, pH 8.0), at a rate of 1 ml/min. The fractions were assayed for citrate synthase activity, as described 2.3.4, and those containing the peak of enzymic activity were pooled.

To remove the oxaloacetate and coenzyme A from the solution of purified citrate synthase, the protein sample was passed through a 35 ml gel filtration column of G-25 Superose. The sample was loaded as a 1.5 ml volume containing 20-30 mg protein, concentrated using an Amicon Ultrafiltration unit with a 10 kDa-cutoff Ultrafiltration membrane. Protein was eluted with 2TE8 buffer at a rate of 0.3 ml/min. Again, the fractions containing the peak of citrate synthase activity were pooled and it was this sample that was used in further analytical studies, unless otherwise stated.

CHAPTER 3

STRUCTURAL ANALYSIS OF CITRATE SYNTHASE FROM *TP. ACIDOPHILUM*

3.1 INTRODUCTION

In order to conduct a comprehensive investigation into the thermostability of any protein, comparative analyses between mesophilic and thermophilic homologs are important; any theories conceived can then be tested by site-directed mutagenesis. Attempts to gain an insight into protein thermostability have been carried out previously using information from amino acid sequences alone [Menendez-Arias & Argos, 1989]; however, these results can be treated as guide lines only, since primary sequences fail to illustrate fully the role played by residues in the context of the 3D structure of a protein. To overcome this obstacle, mesophilic and thermophilic 3D structures should be compared, and from such comparisons rational site-directed mutants can be designed.

The aim of the work in this thesis is to characterise the thermostability of citrate synthase. At the onset of the research, the only available 3D structures of the enzyme were from the mesophilic organisms, pig and chicken [for examples see Remington *et al.* (1982)]. Protein crystals of a thermostable citrate synthase, that from the thermophilic Archaeon *Tp. acidophilum*, had been grown and the structural determination was about to begin. Aware of the lengthy process that this determination could be, it was decided that a 3D structure of the archaeal enzyme would be created by homology modelling, employing the crystal structure of pig citrate synthase as a template. The applications of the protein model would be two-fold: First, it would be used in comparative analyses with the 3D structure of pig citrate synthase, in order to identify features that may play a role in the thermostability of the archaeal enzyme. Secondly, it would be compared to the

crystal structure of *Tp. acidophilum* citrate synthase, once this had been solved, so that a guide could be compiled to aid future modelling of the protein; this is based on the fact that amino acid sequences are determined at a faster rate than 3D structures of proteins are elucidated. However, the crystal structure of the archaeal citrate synthase was solved rapidly [Russell, et al., 1994], and so the model was not required in the comparative analyses with the mesophilic enzyme. In addition, the 3D structure of citrate synthase from the hyperthermophilic Archaeon *P. furiosus*, was recently determined [Russell et al., in press]. From a comparison of the enzymes from pig and the two Archaea, features have been identified that may be important for thermostability of citrate synthase, and these are discussed in section 3.3.2.

3.2 RESULTS

3.2.1 MULTIPLE SEQUENCE ALIGNMENTS

The available citrate synthase amino acid sequences were multiply aligned, by progressive, pairwise comparisons using the Wisconsin GCG Sequences Analysis software package [Devereux et al., 1984] (Fig. 3.1). Gap weights and gap-length weights were varied in order to find the most suitable alignment, criteria for which included conservation of the characterised active site residues. The default values for the two weights (3.0 and 0.1, respectively) were used ultimately. Using the same package, a table of percentage identities between individual sequences was compiled using the relative positions of the sequences in the multiple alignment (Table 3.1). As can be seen from Table 3.1, the identity between the citrate synthases from pig and *Tp. acidophilum* is very low at only 21 %.

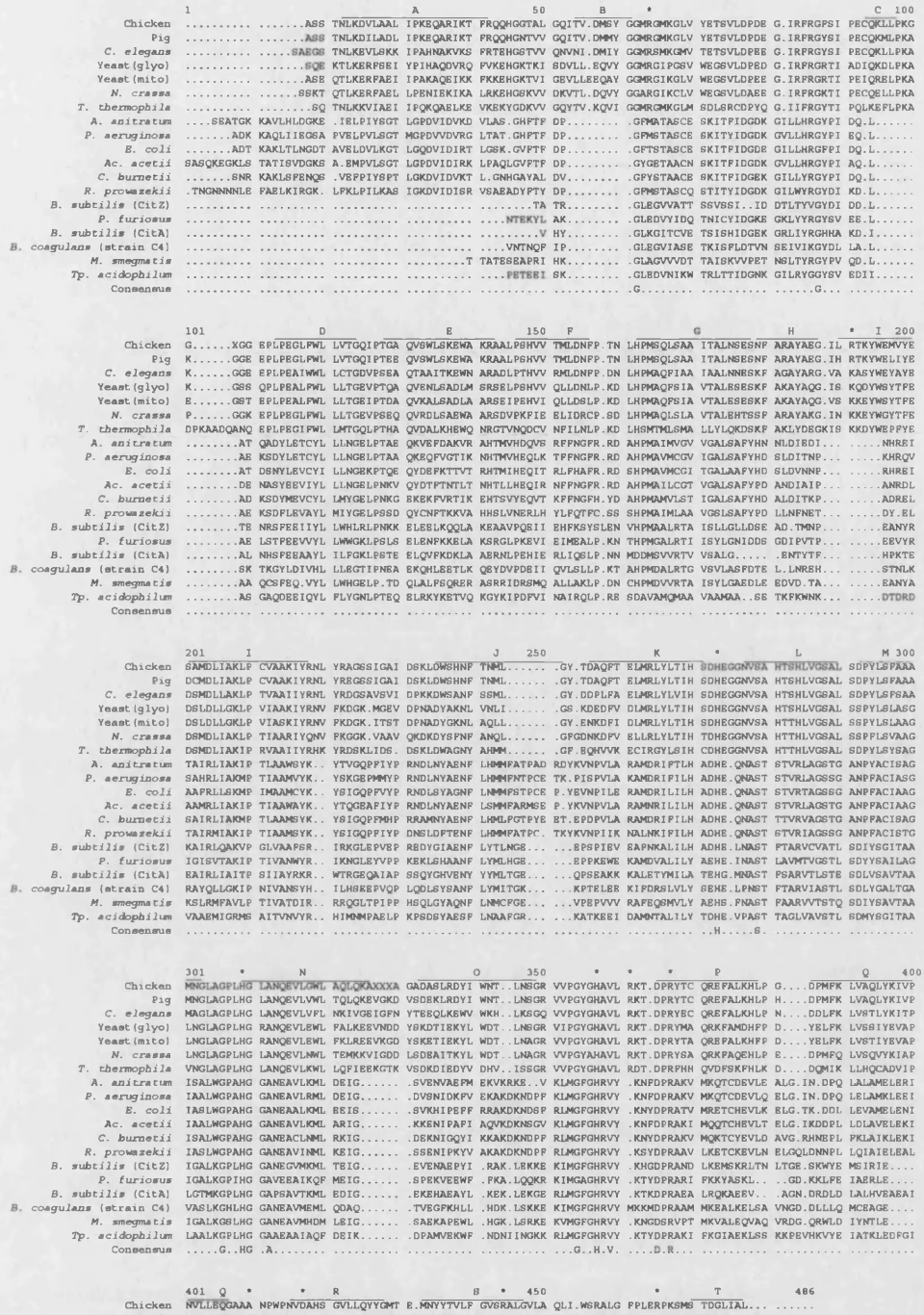


Fig. 3.1 Multiple alignment of all known citrate synthase amino acid sequences. The residues that are conserved throughout all sequences are shown below as the consensus, and the active site residues in the pig enzyme are marked with an *. The α-helical secondary structure of pig citrate synthase is labelled above the alignment (A to T).

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Chicken	91.9	68.6	59.1	56.6	60.1	53.1	21.0	21.7	22.0	22.1	24.3	21.9	24.7	25.3	22.7	21.9	22.0	20.3
2. Pig	100	68.3	58.6	55.9	60.1	51.5	22.5	21.5	22.5	22.6	24.5	23.2	25.8	26.6	23.5	23.0	22.8	20.6
3. *C. elegans*		100	55.8	53.7	54.5	46.5	23.4	23.6	21.8	23.5	25.5	22.7	24.5	27.1	25.1	22.7	23.3	20.6
4. Yeast (glyo)			100	78.7	65.9	47.4	22.7	21.5	23.4	20.2	23.9	22.5	22.9	24.2	21.3	23.5	23.9	21.6
5. Yeast (mito)				100	61.7	45.7	22.2	22.4	22.3	21.6	23.4	22.0	23.4	24.2	21.9	23.2	24.7	22.1
6. *N. crassa*					100	47.4	24.1	23.6	24.1	23.0	26.2	24.8	23.1	24.2	20.8	25.7	22.0	21.1
7. *T. thermophila*						100	23.4	22.2	21.6	21.6	24.8	22.7	23.7	23.9	21.1	21.3	21.5	20.1
8. *A. anitratum*							100	74.9	66.7	64.3	68.3	60.3	34.7	32.7	31.2	34.4	32.2	29.7
9. *Ps. aeruginosa*								100	68.4	66.6	66.8	59.6	34.1	30.9	28.8	31.2	31.1	29.4
10. *E. coli*									100	61.1	63.7	59.5	34.1	33.5	28.5	30.3	30.8	29.7
11. *C. burnetii*										100	59.8	55.6	37.1	33.2	30.1	32.8	30.3	31.3
12. *Ac. acetii*											100	56.9	36.6	32.7	30.9	34.2	31.1	31.0
13. *R. prowazekii*												100	36.0	33.2	29.9	34.7	29.5	31.0
14. *B. subtilis* (CitZ)													100	44.6	41.1	40.2	40.9	35.2
15. *P. furiosus*														100	39.5	36.3	35.7	41.2
16. *M. smegmatis*															100	38.3	34.1	30.9
17. *B. subtilis* (CitA)																100	35.0	30.3
18. *B. coagulans*																	100	30.8
19. *Tp. acidophilum*																		100

Table 3.1 Percentage identities between the amino acid sequences of citrate synthase

A multiple alignment of dimeric citrate synthase amino acid sequences was generated (Fig. 3.2). Sequences from *Caenorhabditis elegans, Neurospora crassa, P. furiosus* and two *Bacillus* species were not included since they were determined after the date of the compilation used in the subsequent homology-modelling exercise. The alignment of dimeric citrate synthases was compared to another alignment of all the enzyme sequences available at the time, again not including those sequences mentioned above. Both alignments were in close agreement, and the similarities described in section 3.2.2.1 were maintained in the alignment of dimeric citrate synthases. It was decided that the multiple alignment of the dimeric citrate synthases would be used in the homology modelling of the enzyme from *Tp. acidophilum*, thus eradicating any bias in the alignment due to the inclusion of hexameric citrate synthases. The validity of the alignment in Fig. 3.2 will be assessed in section 3.2.6.

3.2.2 SEQUENCE COMPARISONS

The multiple alignment of all citrate synthase amino acid sequences (Fig. 3.1) can be used to highlight similarities and differences amongst individual sequences.

3.2.2.1 SEQUENCE SIMILARITIES

The 11 active site residues are particularly well conserved in the alignment, with six of these residues conserved in all 19 sequences. There is a total of 20 residues that are conserved throughout all the sequences (identified as the consensus in Fig. 3.1). Six of these conserved residues are Gly and ten are polar. Using the 3D structure of pig citrate synthase it would appear that 16 of the conserved residues occur within the loop regions between helices, or at the termini of helices. The majority of the polar residues are positioned around the active site residues, and thus it would appear that most of the conserved residues could play an important role in maintaining the structure and function of a fully-active citrate synthase.

Fig. 3.2 Multiple alignment of dimeric citrate synthase amino acid sequences available at the time. The active site residues (*) and the residue number (in [ ]) of the pig enzyme are shown, as are the conserved residues (consensus sequence).

3.2.1.2 SEQUENCE DIFFERENCES

There are three major differences between the citrate synthase sequences, detected from the multiple alignment (Fig. 3.1). A number of the enzyme sequences are shorter than the rest by approximately 40 residues at the N-terminal end of the protein. These 'short-form' enzymes are found in all the archaeal organisms, in the *Bacillus* species and in *Mycobacterium*. The N-terminal peptides of citrate synthase isolated from two Archaea, *Haloferax volcanii* and *Sulfolobus solfataricus* have been determined [James *et al.*, 1991; Lill *et al.*, 1992], and it would seem that these enzymes are also of the 'short-form', since the peptides align with the N-terminus of *Tp. acidophilum* citrate synthase. In pig citrate synthase, the N-terminus of the protein forms a large helix that sits on the surface of the molecule exposed to solvent, and interacts with the same residues of the other subunit upon dimer formation.

The second difference involves four residues that are conserved throughout all the citrate synthase sequences except that from *Tp. acidophilum* (Table 3.2). Two of these residues, Val 96 and Met 106, are found in helix G, which sits at the subunit interface of the *Tp. acidophilum* citrate synthase dimer, although they are not involved in inter-subunit contacts. The other two residues, Ser 132 and Ser 159, are buried within the monomer structure. In pig citrate synthase, Pro 183, the equivalent residue to Ser 132 in the archaeal enzyme, is known to cause the kink in helix I; this helix is still kinked, but not to such a large extent, in the archaeal crystal structure. Ser 159 is involved in six hydrogen bonds through the side chain and main chain, whereas the equivalent residue in pig citrate synthase, Asn 212, has two main chain hydrogen bonds only.

The final obvious difference as seen from the multiple sequence alignment in Fig. 3.1, is the Ala content of helix G. This region of the *Tp. acidophilum* citrate synthase sequence has a significantly higher number of Ala residues (50% of the residues), when compared to the same region of all other sequences, for example, in pig citrate synthase only 20% of these residues are Ala.

Conserved residue in all CS's (pig numbering)	Unique residue in *Tp. acidophilum* CS
Met 138	Val 96
Leu 148	Met 106
Pro 183	Ser 132
Asn 212	Ser 159

Table 3.2 The four residues that are unique to citrate synthase (CS) from *Tp. acidophilum*, but conserved in all the other citrate synthase sequences (represented by the pig citrate synthase residues).

3.2.3 AMINO ACID CONTENT AND EXCHANGES

Using the structural sequence alignment of citrate synthase from pig and *Tp. acidophilum* (Fig. 3.6(a), section 3.2.6), a matrix of amino acid exchanges between the two proteins has been produced (Table 3.3). The amino acid content of both sequences is also shown on the matrix. There are a number of interesting points to note from Table 3.3. For example, the most prominent amino acid type in *Tp. acidophilum* citrate synthase is Ala, whereas in pig citrate synthase it is Leu. The majority of residue types exchanging with the Ala are Asn, Ser and Val. Focusing on the thermophilic protein, there are significantly more Glu and Lys residues, and fewer His, Asn and Gln residues than in the mesophilic protein. The number of Asp and Arg residues is approximately the same in both citrate synthases. The polar amino acids together with Gly, Ala, Leu and Pro are amongst the most conserved residue types between pig and *Tp. acidophilum* citrate synthases.

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	
A	**8**	2	2	1		2		1	1	1	3	4	2			5	2	7	1	2	44
C																					0
D			**5**	4		2	1	1	2			1	1			1	1				19
E	1		1	**3**	1	1		1	3	2		1	1	5	2	2	1	1	1	1	30
F	1				**1**		2	2		5	1	1		1						1	15
G	1			2		**12**			0	1		1	1		2	3		1			26
H			1				**3**					1		1							6
I		2	0	1	2	2		**2**		9	1		2	1			3	3		2	31
K	3		2		1	2		1	**3**	1		1	1	2	3	1	2	1	2		30
L				1	2	1	2	3		**8**	2						1	4			24
M						1				3	**1**		2	2		1		1		1	12
N				1					2			**3**		1	1		1		1	1	15
P	1		1			1			2	1		1	**7**			1	1				17
Q	1			1	2	1				3		1							1		10
R	1		1	1					2	1	1		1		**5**	1	1	1		2	19
S				1			1			5	1		2			**6**				2	19
T	6			1			1	1	1	3		1				2	**2**	2		1	21
V	2		2			2	1	3		3	4				1	1		**2**	1		22
W																		2		1	4
Y	2		1		2		1			3						2	2		2	**4**	20
	33	4	21	24	12	33	14	19	25	53	15	18	22	17	19	29	23	28	9	19	

Table 3.3 Amino acid exchanges between pig citrate synthase (listed horizontally) and *Tp. acidophilum* citrate synthase (listed vertically) using data from the sequence alignment Fig. 3.6(a) Example: 4 Cys residues in pig citrate synthase are exchanged with 2 Ala and 2 Ile residues in *Tp. acidophilum* citrate synthase. The amino acid contents are listed at the sides of the matrix.

3.2.4 SECONDARY STRUCTURE PREDICTIONS

Prior to homology modelling the 3D structure of *Tp. acidophilum* citrate synthase, the secondary structure of the molecule was predicted, from the amino acid sequence, using the program PREDICT [Eliopoulos *et al.,* 1982]. To test the validity of the program, the secondary structure of pig citrate synthase was predicted also and compared with the previously characterised structure [Remington *et al.*, 1982] (Fig. 3.3). PREDICT estimates secondary structures by using the joint prediction from eight individual algorithms [referenced in Eliopoulos *et al.*, 1982]. The propensities of each residue to form part of an α-helix, a β-strand and a turn are calculated and plotted for each algorithm; the consensus is then taken as the predicted structure.

From the prediction program, circular dichroism work [Russell, 1994] and the multiple sequence alignment (Fig. 3.2), it appears that the secondary structure between citrate synthase from pig and *Tp. acidophilum* is quite well conserved.

When the known and predicted secondary structures of pig citrate synthase were compared, it was obvious that the program was fairly good at predicting the **areas** of secondary structure but not accurate enough to predict the **length** of the helices. Since the pattern of predicted structures for *Tp. acidophilum* citrate synthase is similar to that for pig citrate synthase, one can assume that the areas of secondary structure had been defined but the length of helices remained unknown. The secondary structure of *Tp. acidophilum* citrate synthase predicted by the program gave similar results to that predicted from the alignment utilising the known secondary structure of the pig enzyme. This means that the alignment of dimeric citrate synthases should be suitable for use in the modelling.

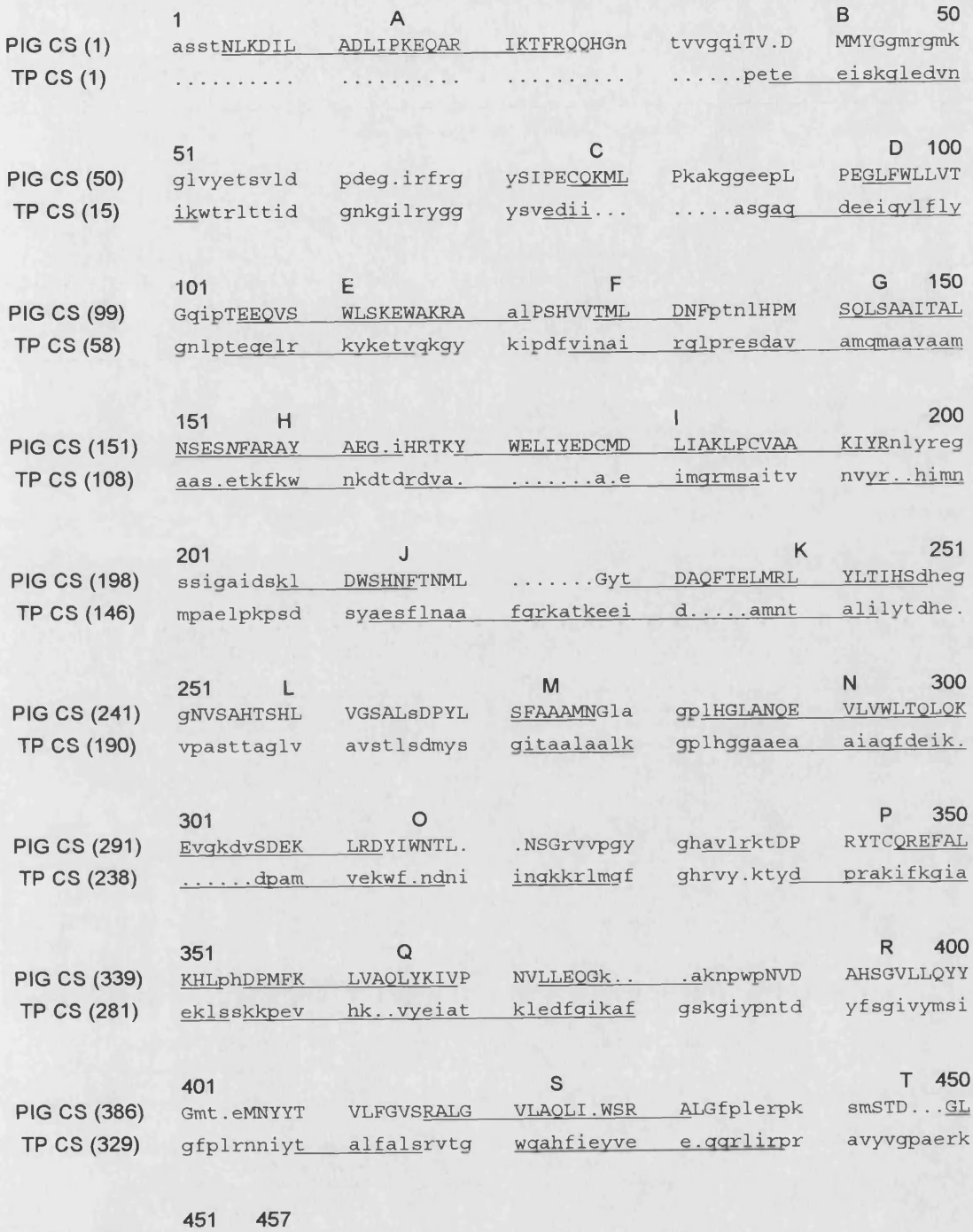


Fig. 3.3 Predicted helical secondary structures (underlined) for citrate synthase (CS) from pig and *Tp. acidophilum*. The residues involved in the actual secondary structure of pig citrate synthase, as determined from the X-ray crystal structure [Remington *et al.*, 1982], are in upper case. The number in brackets refers to the residue number of the amino acid following immediately.

3.2.5 HOMOLOGY MODELLING OF *TP. ACIDOPHILUM* CITRATE SYNTHASE

The 3D structure of *Tp. acidophilum* citrate synthase was homology modelled using the crystal structure of the open form of the pig heart enzyme as the template (2CTS.PDB) (section 2.1). The multiple sequence alignment of dimeric citrate synthases was used to define the residue replacements, insertions and deletions from pig to *Tp. acidophilum* citrate synthase. Analysis identified five regions of the model where the residues have low suitability to their environment. These regions corresponded to the N and C-termini, helix H, the region prior to and including helix K, and finally the C-terminal end of helix Q. The conformation of the sequences of residues involved in these regions could not be altered any further without drastically changing the secondary structure of the protein molecule.

Comparing the 3D structure of pig citrate synthase with that of the modelled archaeal enzyme, it was clear that the overall 3D structure was very well conserved (Fig. 3.4) with an rms deviation of 1.85 Å over the entire peptide backbone. The secondary structure was also well conserved (Fig. 3.5), as suggested by the predictions in section 3.2.4. Previous work has estimated the error margin of a protein model compared to the actual structure to be approximately 3.4Å [Bohm & Jaenicke, 1994], and since this level of error is large enough to affect interactions between residues, the interactions within the monomer and between the subunits of the modelled citrate synthase were not investigated.

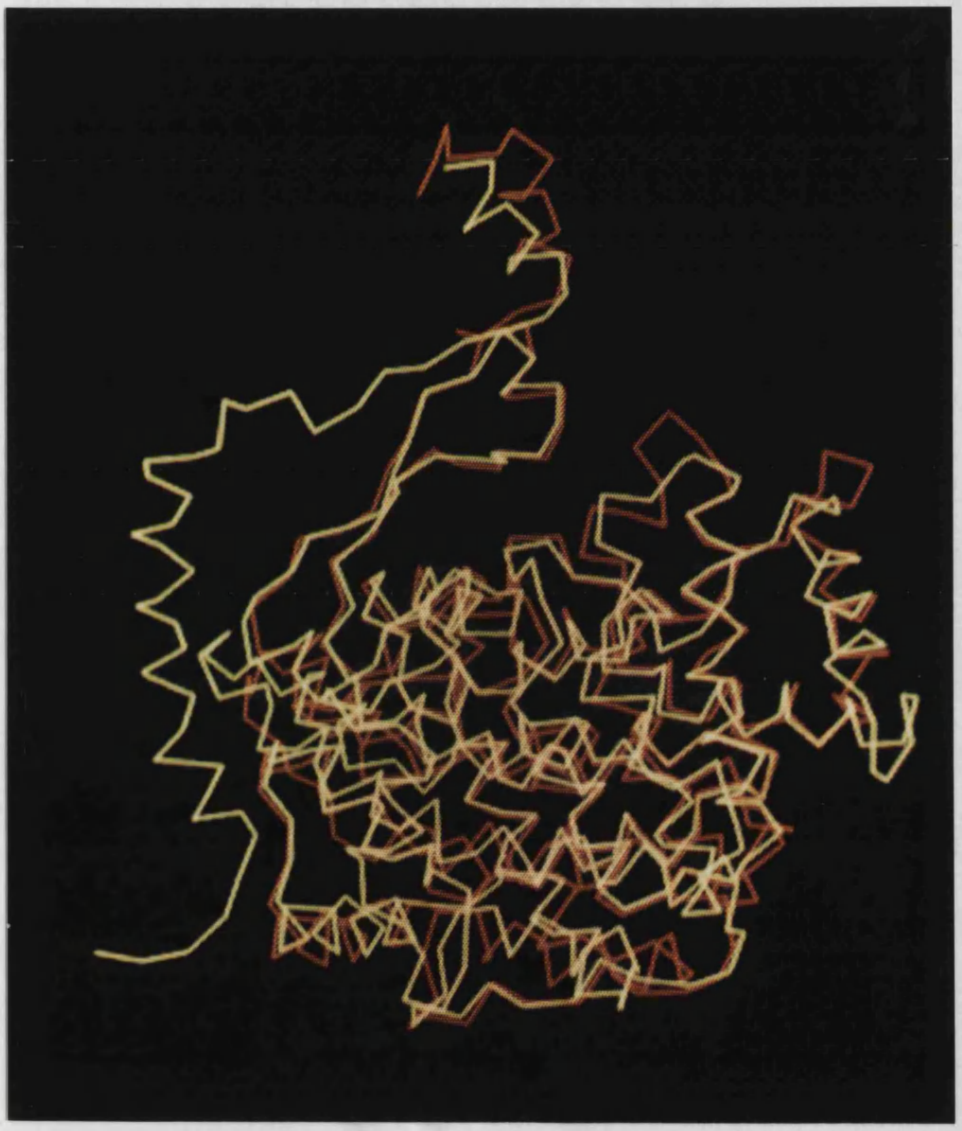


Fig. 3.4 α-Carbon traces of the crystal structure of the pig citrate synthase subunit (yellow) and the modelled *Tp. acidophilum* citrate synthase subunit (red).

```
                1                A                              B          50
PIG CS (1)      asstNLKDIL  ADLIPKEQAR  IKTFRQQHGn  tvvgqiTVDM  MYGgmrgmkg

TP CS (1)       ..........  ..........  ..........  .....PETEE  ISKgledvni

               51                              C                  D  100
PIG CS (51)     lvyetsvldp  deg.irfrgy  SIPECQKMLp  kakggeepLP  EGLFWLLVTG

TP CS (16)      kwtrlttidg  nkgilryggy  svedii....  ....asgaqD  EEIQYLFLYg

               101          E                  F                  G  150
PIG CS (100)    qipTEEQVSW  LSKEWAKRAa  lPSHVVTMLD  NFptnlHPMS  QLSAAITALN

TP CS (58)      nlpTEQELRK  YKETVQKGyk  iPDFVINAIR  QLpresDAVA  MQMAAVAAMa

               151       H                        I                  200
PIG CS (150)    SESNFARAYA  EGi.HRTKYW  ELIYEDCMDL  IAKLPCVAAK  IYRnlyregs

TP CS (108)     aS.ETKFKWN  KDtd......  rd.va..aEM  IGRMSAITVN  VYR..himnm

               201               J                  K                  250
PIG CS (199)    sigaidsklD  WSHNFTNML.  G.yt.DAQFT  ELMRLYLTIH  SdheggNVSA

TP CS (146)     paelpkpsdS  YAESFLNAAF  GRKAtk.eEI  DAMNTALILY  TdheV.PAST

               251  L               M                       N          300
PIG CS (246)    HTSHLVGSAL  sDPYLSFAAA  MNGlagplHG  LANQEVLVWL  TQLQKEvgkd

TP CS (195)     TAGLVAVSTL  sDMYSGITAA  LAALKgplhg  GAAEAAIAQF  DEI...k...

               301       O                                 P          350
PIG CS (296)    vSDEKLRDYI  WNTLN.SGrv  vpgyghavlr  ktDPRYTCQR  EFALKHLphD

TP CS (241)     DP.AMVEKWF  NDNIINGkkr  lmgfghrvy.  ktyDPRAKIF  KGIAEKlssK

               351       Q                       R                  400
PIG CS (345)    PMFKLVAQLY  KIVPNVLLEQ  Gk..aknpwp  NVDAHSGVLL  QYYGmt.eMN

TP CS (286)     KPEVHK.VYE  IATKLEDFGI  KAfgskgiyp  ntDYFSGIVY  MSIGfplRNN

               401            S                        T          450
PIG CS (392)    YYTVLFGVSR  ALGVLAQLIW  SRALGfpler  pksmSTDGLI  KLv..d..sk

TP CS (335)     IYTALFALSR  VTGWQAHFIE  YVEEQqrlir  pravYVGPAE  RKYVPIaerk
```

Fig. 3.5 Secondary structures (in upper case) in the 3D model of *Tp. acidophilum* citrate synthase are compared to those of pig citrate synthase. The helices of the pig enzyme are labelled A to T. The alignment is a structural alignment compiled using the Structure Homology Program [Stuart, 1979].

3.2.6 ANALYSING THE VALIDITY OF HOMOLOGY MODELLING

Shortly after the model of *Tp. acidophilum* citrate synthase was created, the crystal structure of the protein was solved to 2.5 Å resolution [Russell *et al.*, 1994]. Understandably, the crystal structure of the archaeal enzyme was to be used in further comparative analyses between mesophilic and thermophilic citrate synthases. At this point, the validity of the model was assessed by a comparison with the X-ray crystal structure from which the weak features of the modelling process can be identified and taken into account in future modelling studies. The three aspects of modelling that were analysed included the sequence alignment, the secondary structure and the overall conformation of the protein molecule.

(a) The first step in modelling a protein involves compiling a suitable alignment of the amino acid sequence of the template protein with that of the protein to be modelled. The sequence alignment used in the modelling procedure was that of dimeric citrate synthases (Fig. 3.2). During the process of homology modelling, particular areas of the basic alignment were not adhered to strictly since some changes were very difficult to accommodate into the 3D template structure of the pig enzyme. Therefore, for the comparisons, a structural sequence alignment was generated, as shown in Fig. 3.5, using the Structure Homology Program [Stuart, 1979] to superimpose the X-ray crystal structure of citrate synthase from pig on both the modelled and the crystal structures of the archaeal enzyme (Fig. 3.6). Using Fig. 3.6, the relative alignment of the two *Tp. acidophilum* citrate synthase sequences was analysed. Overall, the archaeal sequences appear to be quite well aligned with a 69% identity between the residues. In the majority of cases, the areas that are not in agreement occur in the loop regions between helices, or at the termini of helices such as H, I, K and Q.

```
                1                   A                                   B                  50
PIG CS (1)      asstNLKDIL  ADLIPKEQAR  IKTFRQQHGn  tvvgqiTVDM  MYGgmrgmkg
a) TP CS (1)    ..........  ..........  ..........  ...PETEE..  Iskgledvni
b) TP CS (1)    ..........  ..........  ..........  .....petee  iskgledvni

               51                                   C                  D100
PIG CS (51)     lvyetsvldp  .deg.irfrg  ySIPECQKML  pkakggeepL  PEGLFWLLVT
a) TP CS (16)   kwtrlttidg  nkgi.lrygg  ysvediias.  .g......aq  deeiqylfly
b) TP CS (16)   kwtrlttidg  .nkgilrygg  ysvedii...  .....asgaq  deeiqylfly

              101           E                     F                  G150
PIG CS (99)     GqipTEEQVS  WLSKEWAKRA  alPSHVVTML  DNFptnlHPM  SQLSAAITAL
a) TP CS (57)   gnlpteqelr  kyketvqkgy  kipdfvinai  rqlpresdav  amqmaavaam
b) TP CS (57)   gnlpteqelr  kyketvqkgy  kipdfvinai  rqlpresdav  amqmaavaam

              151           H                     I                   200
PIG CS (149)    NSESNFARAY  AEGi.HRTK.  YWELIYEDCM  DLIAKLPCVA  AKIYRnlyre
a) TP CS (107)  aasetkf.k.  .........w  nkdtdrdvaa  emigrmsait  vnvyrhim.n
b) TP CS (107)  aas.etkfkw  nkdtd.....  ..rd.va..a  emigrmsait  vnvyr..him

              201                   J                   K                   250
PIG CS (197)    gssigaidsk  lDWSHNFTNM  L.G.yt.DA.  QFTELMRLYL  TIHSdheggN
a) TP CS (145)  mpaelpkpsd  syaesflnaa  f.g.rk.atk  eeidamntal  ilytdhev.p
b) TP CS (144)  nmpaelpkps  dsyaesflna  afgrkatk..  eeidamntal  ilytdhev.p

              251  L                M                          N          300
PIG CS (243)    VSAHTSHLVG  SALsDPYLSF  AAAMNGlagp  lHGLANQEVL  VWLTQLQKEv
a) TP CS (192)  asttaglvav  stlsdmysgi  taalaalkgp  lhggaaeaai  aqfdeik...
b) TP CS (192)  asttaglvav  stlsdmysgi  taalaalkgp  lhggaaeaai  aqfdei...k

              301           O                                       P350
PIG CS (293)    gkdvSDEKLR  DYIWN.TLN.  SGrvvpgygh  avlrktDPRY  TCQREFALKH
a) TP CS (241)  dpam....ve  kwfndniin.  gkkrlmgfgh  rvyktydpra  kifkgiaekl
b) TP CS (241)  ...dp.amve  kwfnd.niin  gkkrlmgfgh  rvy.ktydpr  akifkgiaek

              351                   Q                                   R400
PIG CS (341)    Lp....hDPM  FKLVAQLYKI  VPNVL.LEQG  k..aknpwpN  VDAHSGVLLQ
a) TP CS (283)  sskkpevhkv  yeiatkledf  gikafgs...  ....kgiypn  tdyfsgivym
b) TP CS (282)  ls....skkp  evhk.vyeia  tkled.fgik  afgskgiypn  tdyfsgivym

              401                   S                                   T450
PIG CS (384)    YYGmt.eMN.  YYTVLFGVSR  ALGVLAQLIW  SRALGfpler  pksmSTDGLI
a) TP CS (326)  sigfp.lrnn  iytalfalsr  vtgwqahfie  yveeqqrlir  pravy.vGPA
b) TP CS (326)  sigfplrnn.  iytalfalsr  vtgwqahfie  yveeqqrlir  pravyvgpae

              451           465
PIG CS (432)    KLv..d..sk  .....
a) TP CS (375)  ERK..Y..VP  IAERK
b) TP CS (375)  rkyvpiaerk  .....
```

Fig. 3.6 Structural alignment of citrate synthases generated from the X-ray crystal structures of the enzyme from pig, *Tp. acidophilum* (a) and the 3D model of the latter protein (b). The 3D structure was not designated for certain residues at the N- and C-termini of the crystal structure of *Tp. acidophilum* citrate synthase (shown in upper case) due to poor electron density in these areas. The α-helices of pig citrate synthase are shown in bold uppercase and are labelled A to T.

(b) The secondary structures found in the model and crystal structure of *Tp. acidophilum* citrate synthase were compared. The positions of the structures are shown in Fig. 3.7, and the rms deviations over all the main chain atoms of the helices are shown in Table 3.4. There is a high degree of homology between the majority of the secondary structures as indicated by the low rms deviation values. The areas of lower agreement include the region around helices I, P and Q and also the termini of many helices. A major difference between the secondary structures involves Helix H, a helix that is present in the crystal structure of pig citrate synthase and the model of *Tp. acidophilum* citrate synthase, but is absent in the crystal structure of the latter protein.

```
                                                    C
  1  peteeiskgl  edvnikwtrl  ttidgnkgil  ryggysVEDI  IAsgaqDEEI  50

       D             E                      F             G
 51  QYLFLYgnlp  tEQELRKYKE  TVqkqykipD  FVINAIRqlp  resdAVAMQM  100

                                         I
101  AAVAAMAASe  tkfkwnkdtD  RDVAAEMIGR  MSAITVNVYR  HImnmpaelp  150

                J                        K                L
151  kpsdsYAESF  LNAAFqrkat  KEEIDAMNTA  LILYtdhevp  ASTTAGLVAV  200

                M                        N                O
201  stlsDMYSGI  TAALAALkqp  lhggAAEAAI  AQFDEIkdpa  mVEKWFNDNi  250

                                    P                Q
251  ingkkrlmgf  ghrvyktydP  RAKIFKGIAE  KLSskKPEVH  KVYEIATKLE  300

                               R                S
301  DFGIKAFgsk  giypnTDYFS  GIVYMSIqfp  lrnnIYTALF  ALSRVTGWQA  350

351  HFIEYVEEqq  rlirpravyv  qpaerkyvpi  aerk  384
```

Fig. 3.7 Helical secondary structures of *Tp. acidophilum* citrate synthase as determined from the X-ray crystal structure [Russell *et al.*, 1994] (upper case and with helices labelled C to S) and from the modelled protein (underlined). The secondary structures of residues at the N- and C-termini of the crystal structure (italics) have not been determined due to poor electron density in these areas.

Helix	RMS deviation (Å)	Helix	RMS deviation (Å)
C	2.64	L	1.27
D	1.97	M	0.88
E	1.74	N	2.05
F	1.92	O	2.86
G	0.84	P	4.94
I	3.18	Q	5.51
J	2.42	R	2.17
K	1.25	S	1.90

Table 3.4 RMS deviations between the α-helices of the crystal structure of *Tp. acidophilum* citrate synthase [Russell *et al.*, 1994] and those of the modelled protein. Values were calculated using the program ASH [Stuart *et al.*, 1979] to superimpose the two structures upon each other. Since the model is of the closed citrate synthase form and the crystal structure is of the open citrate synthase form, the superimposition was carried out in two stages, determined by the residues of the small and large domains, residues 225-327 being the former. The secondary structure in the X-ray crystal structure was used to define the residues of each helix. The secondary structures of the N- and C-termini of the crystal structure have not been determined due to poor electron density in these areas.

As well as comparing the helical structures of the protein, the loop regions joining such structures have also been compared. There are particular loops in pig citrate synthase that are significantly longer than those found in the crystal structure of the archaeal enzyme; for example, those between helices G and I and between helices J and K [Russell *et al.*, 1994]. If these loops are compared to the same regions of the modelled *Tp. acidophilum* citrate synthase, it is clear that the modelled loops are of an intermediate size between those in the mesophilic and thermophilic crystal structures (Fig. 3.8).

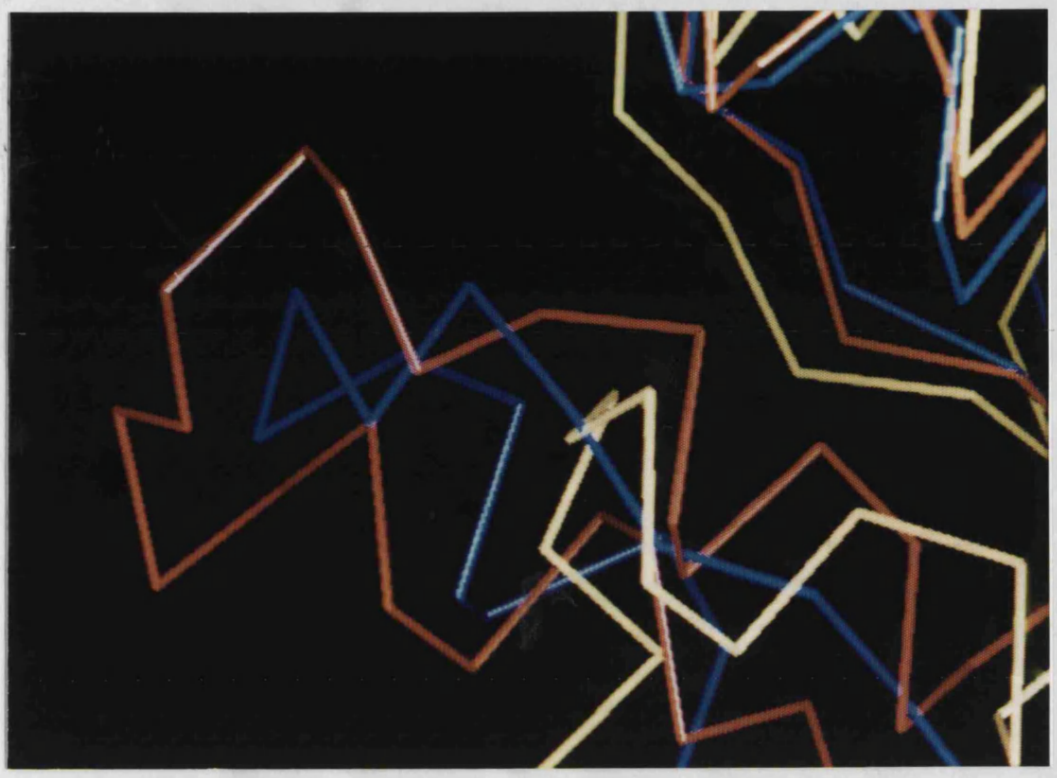


Fig. 3.8 Comparison of loop sizes in the structures of citrate synthase from pig (red), and *Tp. acidophilum*. For the latter protein, the model is coloured blue and the X-ray crystal structure is in yellow. Created using O [Jones *et al.*, 1991].

Between the actual and modelled structures of *Tp. acidophilum* citrate synthase, the rms deviation values vary considerably for the main chain atoms of the regions of structure not listed in Table 3.4. For most residues, the values range from 0.6-4.0 Å however, for the N- and C- termini, the region between Helices G and I (Helix H in the model) and the loop between Helices J and K, the rms deviations are excessively high at values up to 13.5 Å.

A comparison of the overall conformation of the crystal structures of pig and *Tp. acidophilum* citrate synthases, and of the modelled structure of the latter protein, are shown below (Table 3.5). From these results it would appear that the accessible surface area of the model is intermediate between the crystal structures

of the mesophilic and thermophilic enzymes, and that the volume of the model is in the region that would be expected for the closed conformation of *Tp. acidophilum* citrate synthase. The percentage of buried atoms in the modelled structure is noticeably less than those values in the three crystal structures. It is expected that the level of buried atoms would increase upon substrate binding, as happens to the degree of 3.8% in the pig enzyme, but from a comparison of the values in both structures of *Tp. acidophilum* citrate synthase, it appears that the percentage of atoms buried actually decreases in this protein. This decrease is likely to be related to the low level of buried atoms in the modelled structure which in turn is probably due to the sub-optimal packing of atoms.

CS STRUCTURE	RMS DEVIATION (Å)	ACCESSIBLE SURFACE AREA ($Å^2$)	VOLUME ($Å^3 \times 10^4$)	ATOMS BURIED (%)
Pig				
(a) closed	1.85 to (c)	32,066	9.99	52.0
(b) open	2.15 to (d)	33,417	9.96	50.1
Tp. acidophilum				
(c) model	3.4 to (d) [large domain] 4.6 to (d) [small domain]	32,080	8.78	39.4
(d) crystal		30,249	8.71	50.7

Table 3.5 **A comparison of the overall conformation of citrate synthase structures.** The rms deviation was calculated over the Cα backbone, using the Structure Homology Program [Stuart, 1979] when involving pig citrate synthase, and the program ASH [Stuart *et al.*, 1979] for the two archaeal citrate synthase structures. The range of residues used for the large domain of *Tp. acidophilum* citrate synthase encompassed 7-224 and 328-370, whilst the residues used to represent the small domain were 225-327. The accessible surface area and the percentage of atoms buried were calculated for the dimer of the three citrate synthase structures, with the program GRASP [Nichols *et al.*, 1991]. The volume, also of the dimeric structures, was calculated using the program VOIDOO [Kleywegt & Jones,1994]

3.3 DISCUSSION

The 3D structure of *Tp. acidophilum* citrate synthase was homology modelled (section 3.2.5), primarily so that it could be used in comparative analyses with the crystal structure of the pig enzyme to identify features that may be important for conferring protein thermostability. At the point that this analysis began, the crystal structure of *Tp. acidophilum* citrate synthase was solved [Russell *et al.*, 1994], and thus, was used from then on in place of the model. A second application of the model was to be for use in comparisons with the solved crystal structure of the same enzyme, in order to assess the validity of the homology modelling technique.

3.3.1 VALIDITY OF HOMOLOGY MODELLING

The sequence alignment and secondary structure comparisons that were carried out between the model and crystal structure of *Tp. acidophilum* citrate synthase (section 3.2.6) show that on the whole the two structures appear well matched. As would be expected, there are areas of disagreement, and these tend to fall into three categories:

(i) Firstly, there are regions where the structure of the modelled citrate synthase favours that of the template protein (for example, with the absence of Helix H); this is due to inaccuracies at the sequence alignment (Fig. 3.6) and secondary structure prediction level (Fig. 3.3).

(ii) There are other areas of the protein (such as the loop regions (Fig. 3.8)), where the model has an intermediate structure between the two crystal structures of citrate synthase; this appears to be the result of inevitable biasing when using the pig citrate synthase structure as a template for the modelling procedure. Hence, when modelling another thermostable/archaeal citrate synthase it would be wiser to use the crystal structure of the *Tp. acidophilum* enzyme, instead of the mesophilic enzyme structure, as a template. Of course, sequence identity also must be taken into account when choosing which template structure to employ, so that the chosen

structure is the one whose amino acid sequence shares the highest identity with that of the protein to be modelled. In the present case the generated model is quite accurate considering that the sequence identity between pig and *Tp. acidophilum* citrate synthase is only 21%.

(iii) The final category involves regions of the protein where the model favours neither crystal structure (such as the definition of the termini of helices and the value for percentage of buried atoms); these disagreements seem to be due to inaccuracies in the secondary structure prediction program, in the former example, and the manual modelling technique employed.

Caution must be exercised when choosing programs that aid the modelling of a protein structure. Compilation of a sequence alignment is the initial step in the modelling technique and it became apparent from the comparative analyses in section 3.2.6 that the sequence alignment and the interpretation of it are indeed crucial steps. Regions within an alignment that incorporate deletions or insertions are obviously the areas of the structure that are the most difficult to model, and it seems that it was these areas in the modelled structure that had the lowest suitability to their environments and also the highest rms deviation from the actual structure of the enzyme. Most noticeable of these areas was the region around Helix H, and the loop between Helices J and K. There are a variety of sequence alignment programs available and the one used in this work, *Pileup* from the GCG Sequence Analysis Software package [Devereux *et al.*, 1984], involves progressive pairwise alignments using the Dayhoff table of amino acid similarities for reference. There are those programs that align sequences biased towards the known secondary structure of one, or more, of the proteins and penalise gaps inserted into secondary structure regions, for example, the *AMPS* program [Barton & Sternberg, 1990]. This type of program is fine for aligning the sequence of proteins that have the same secondary structural pattern, but clearly the program falls down where one of the proteins aligned lacks at least one secondary structural element. For example, Helix H is present in pig citrate synthase but is absent in the enzyme from *Tp. acidophilum*, and thus in a true alignment of these sequences there is a gap at this point in the archaeal sequence (Fig 3.6(a)). Coincidentally, Helix H was not

deleted in the sequence alignment that was used to create the model of citrate synthase (Fig. 3.2) although, during the modelling process the helix did disappear due to additional strain from the deletions following this region; however, it was re-introduced manually ! If the alignment is taken to be optimal for such low sequence identity, it would appear that one method of increasing the similarity between the modelled and the actual structures would be to concentrate on increasing the suitability of the identified residues to their environment. As mentioned previously in section 3.2.5, these residues were not altered any further since this would have required a significant change in the conformation of the zone of residues around that point; however, in the case of Helix H, it seems that that is precisely what was needed.

The prediction of secondary structures is the second step in homology modelling a protein. The programs available to carry out such a task are unlikely to be 100% accurate, as illustrated by the comparison of the predicted structure with the true structure of pig citrate synthase (Fig. 3.3). Indeed, secondary structure prediction programs have only a 70% success rate at best, even with the most sophisticated algorithms. The point at which the prediction fell short was in the assignation of helix termini, although the regions that were predicted as α-helical were quite accurate. The program that was used in section 3.2.4, *PREDICT* from the GCG Sequence Analysis Software package [Devereux *et al.*, 1984], has to be one of the most accurate since it determines a consensus prediction from the results of eight individual algorithms. Having said that, the results of secondary structure prediction programs should only be used as a template and the modeller should build the secondary structure of the protein taking into consideration the interactions between both contiguous and non-contiguous residues.

From the comparison of secondary structures between the modelled and actual structures of *Tp. acidophilum* citrate synthase, helices I, P and Q were identified as having high rms deviations (Fig. 3.8). The discrepancy with Helix I is due to the complications concerning the presence of Helix H; these in turn are the result of inaccuracies in both the sequence alignment, as discussed above, and also in the secondary structure predictions. The high rms deviation between Helices P and Q

is an example of how the modelled structure favours that of the template as opposed to tending towards the true structure of the archaeal enzyme. Helices O, P and Q are part of the small domain in both the mesophilic and thermophilic crystal structures of citrate synthase. In the small domain, these helices have the highest rms deviation between the crystal structures because their orientation is shifted towards a more closed form in the protein from *Tp. acidophilum* compared to pig citrate synthase.

In summary, areas of a modelled protein, especially loop regions between helices, can appear to be a somewhat inaccurate account of the true protein structure. These areas tend to involve residues close to, or at, the positions that incorporate insertions and deletions in the sequence alignment and also have low suitability to their environment. If the sequence identity is low, as is the case here, the frequency of gaps is likely to increase. Taking the secondary structure predictions as only an outline, the troublesome residues should be altered so as to increase the suitability to their environment, making gross secondary structural changes if necessary, as illustrated with Helix H. In doing so, there is a higher probability that the resultant structure is more likely to resemble the true structure, where all the amino acids should be within a stable environment.

For other regions of the protein, most prominently around the active site and in helical structures, there is good agreement between the modelled and actual structures. This, of course, tends to increase as the sequence identity of the template protein and the modelled protein increases. Although the secondary structure and the overall conformation of the modelled protein can be analysed with some degree of confidence, it is still unwise to investigate interactions between residues of the molecule due to the 3.4 Å error margin that has been determined for the homology-modelling technique [Bohm & Jaenicke, 1994], unless the sequence identity is very high. Nevertheless, in the absence of a required 3D structure, a modelled protein suffices adequately until the true structure is solved.

3.3.2 POTENTIAL THERMOSTABILISING FEATURES

The amino acid sequences and the 3D structures of citrate synthases from the mesophilic organism pig, and the moderately thermophilic Archaeon *Tp. acidophilum*, were used in comparative analyses (section 3.2, and Russell [1994]). From these analyses, features have been identified that could play a role in conferring the thermostability upon the archaeal protein, and are discussed below. Various features have then been tested with site-directed mutagenesis in the following chapters.

3.3.2.1 PROTEIN FLEXIBILITY AND COMPACTNESS

Citrate synthase from *Tp. acidophilum* appears to be a more compact molecule than the same enzyme from pig citrate synthase. This can be attributed to the majority of the features that have been identified from the comparative analyses. A more tightly packed structure would account for the rigidity of thermostable proteins at mesophilic temperatures.

(i) Domain Orientation

When superimposing the crystal structures of pig and *Tp. acidophilum* citrate synthase upon each other, a subtle difference can be see in the orientation of the small domain of each protein with respect to their large domain. In the thermostable citrate synthase two of the five helices of the small domain superimpose closely on those of pig citrate synthase, but the remaining three helices are all shifted towards a more closed form of the enzyme. It has been reported that the closed form of pig citrate synthase is more stable to denaturants than the open form, due to the inaccessibility of the protein interior [Srere, 1963]. Working on the same theory, the open form of citrate synthase from *Tp. acidophilum* may be more stable to denaturants than the same form of the pig enzyme, due to the semi-closed nature of the small domain with respect to the large domain. This feature of domain

compactness has also been noted in the comparative analyses of phosphoglycerate kinase [Davies *et al.*, 1993] and thermolysin [Pauptit *et al.*, 1988].

(ii) Helical Deletions

The archaeal citrate synthase possesses fewer helices than the pig enzyme. The most significant of these deletions is the solvent exposed N-terminal helix, but also included is Helix H (of pig citrate synthase), a small buried secondary structure. Both of these deletions ensure that the archaeal protein is more compact. This feature has been investigated with site-directed mutagenesis (Chapter 6) to assess the role that the N-terminal arm plays upon the activity and stability of the protein molecule.

(iii) Loop Deletions

A number of the loops joining the secondary structures within the citrate synthase molecule are shorter in the archaeal enzyme than in the mesophilic protein. This feature has two consequences: making the thermophilic molecule more compact, and possibly increasing the stability of the structure due to the absence of flexible regions that are also accessible to denaturants. From molecular dynamics simulations on bovine pancreatic trypsin inhibitor, it has been shown that the loop regions are the most flexible areas of the protein, and at higher temperatures they are likely to be the first regions to unfold [Daggett & Levitt, 1993]. In contrast, the thermostability of the enzyme β-glycosidase from *S. solfataricus* appears to be the result of a lattice of intramolecular ionic interactions on the surface of the molecule, involving residues within surface loops [I. Sanderson, UKEN meeting, 1995]. Thus it would seem that the two archaeal enzymes, citrate synthase and β-glycosidase, may have achieved the thermostability by different mechanisms.

(iv) Subunit Interface Interactions

It has been proposed that pig citrate synthase undergoes a dimer to monomer transition before monomer unfolding [McEvily & Harrison, 1986], and therefore

differences in the interactions at the subunit interface, between the enzymes from pig and *Tp. acidophilum*, may be important for stability of the latter structure. The subunit interface of the dimer of citrate synthase consists of eight helices, four from one subunit (F, G, M and L) that are opposite and anti-parallel to the same four of the other subunit (F', G', M' and L') (Fig. 3.7).

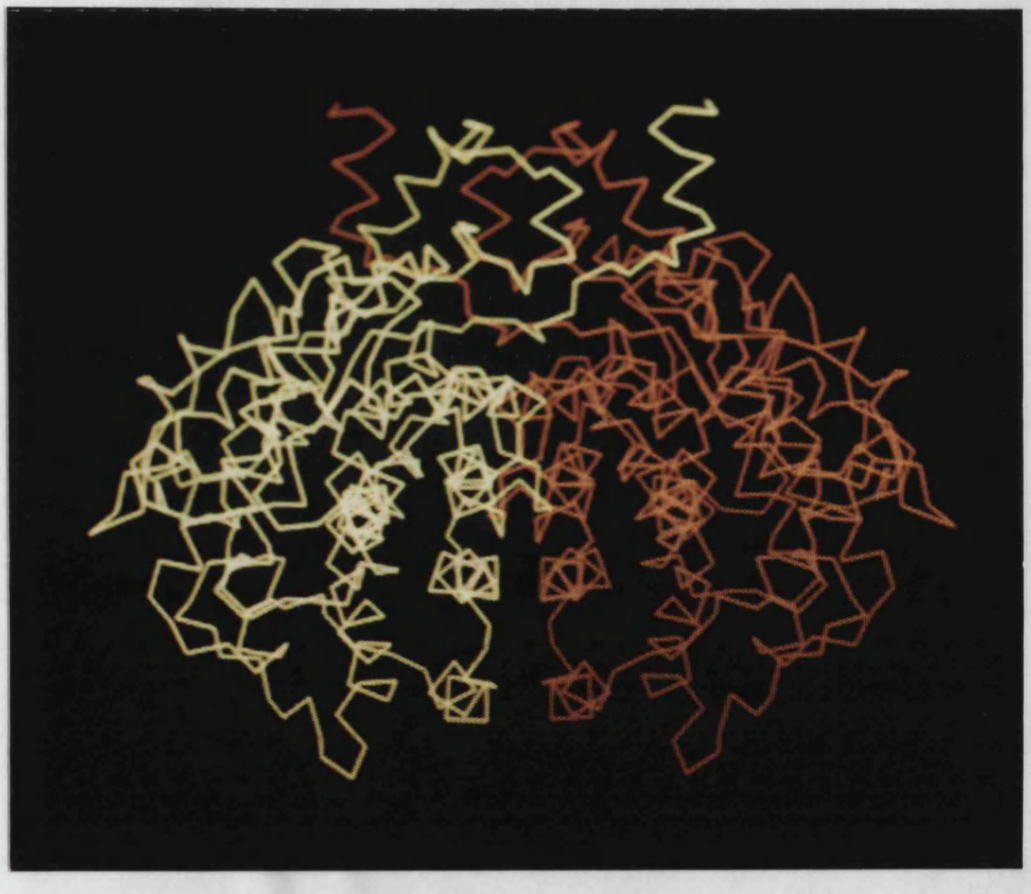


Fig 3.7 The α-carbon trace of the pig citrate synthase dimeric structure, highlighting the four helical pairs at the subunit interface [Remington *et al.*, 1982]. The individual subunits are coloured red or yellow.

In the mesophilic enzyme the interactions between the two subunits involve both hydrogen bonding, between helix pair GG', and hydrophobic bonding, between helix pairs FF', LL' and MM'. Helix G possesses two polar residues, Ser 139 and Thr 146, that are involved in the hydrogen bonding. In *Tp. acidophilum* citrate synthase all

four helices are involved in hydrophobic interactions and the two polar residues have been exchanged for two Ala amino acids, Ala 97 and Ala 104 (Fig. 3.8).

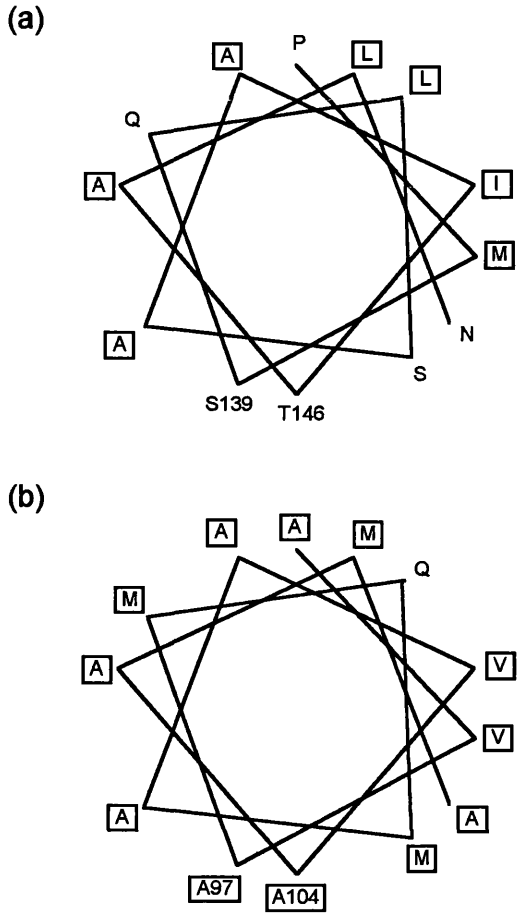


Fig. 3.8 Helical wheels of helix G in citrate synthase from (a) pig and (b) *Tp. acidophilum*. The inter-subunit side of the helix is the lower portion, the hydrophobic residues are boxed, and the residues of interest are identified with their respective residue numbers.

The two exchanges of Ser/Thr to Ala, and the fact that they occur in an α-helix at the subunit interface, are in agreement with the protein-stabilising traffic rules suggested by Argos [Menendez-Arias & Argos, 1989]; this is related to the fact that Ala is the amino acid with the highest helical propensity [O'Neil & Degrado, 1990; Horovitz *et al.*, 1992]. It has also been suggested that the number of salt bridges at the subunit interface of citrate synthase is equivalent in the mesophilic and

thermostable enzymes [Russell *et al.*, in press], possibly indicating that as the stability of the protein rises, hydrophobic interactions at the interface play an increasingly important role in maintaining the integrity of the dimer.

This example of increased hydrophobicity at the interface has been demonstrated in lactate dehydrogenase [Kotik & Zuber, 1993]: following comparisons of the enzymes from *B. megaterium* and *B. stearothermophilus*, a Ser and Thr residue were mutated to Ala amino acids in the mesophilic protein and the result was a 20°C rise in the thermostability of the enzyme. A similar investigation was conducted with pig citrate synthase mutating the two polar residues to Ala (Chapter 4).

(v) Cavities

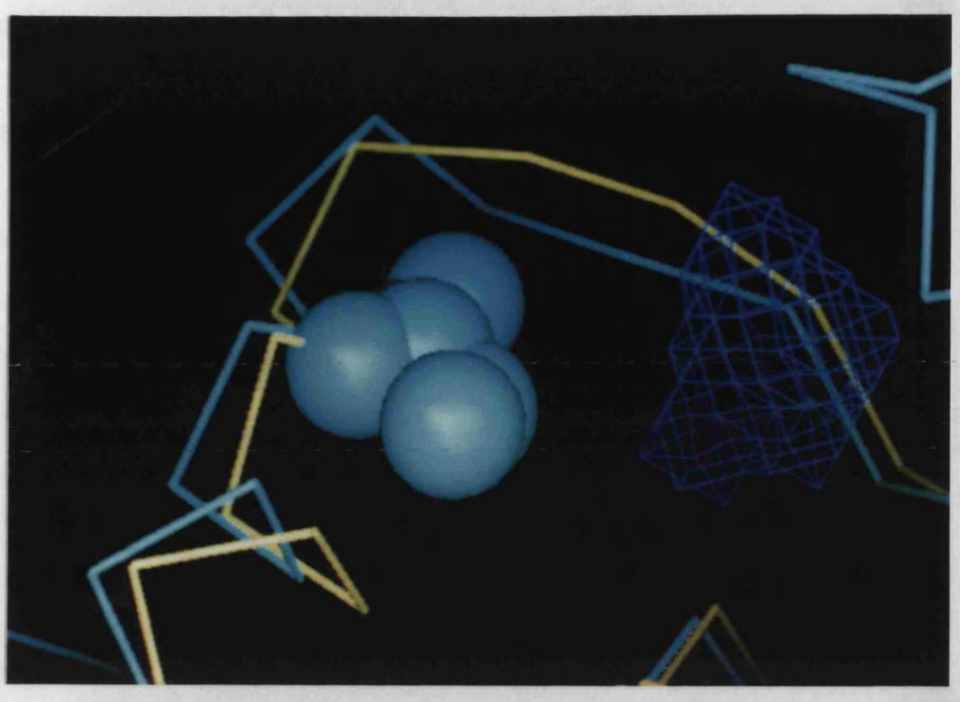
Citrate synthase from *Tp. acidophilum* has a 30-40% smaller volume occupied by cavities compared to the pig enzyme structure, leading to a more compact enzyme (Fig 3.9). The volume of a cavity is defined and thus calculated as the volume which can accommodate a water molecule of 1.4Å radius.

This property has been studied quite extensively by site-directed mutagenesis. Cavity-creating mutants consistently decrease the stability of the protein under investigation [Eriksson *et al.*, 1992], whilst cavity-filling mutants have resulted in both increases and decreases in the stability [Ishikawa *et al.*, 1993(b)]. This feature also has been investigated by site-directed mutagenesis, in Chapter 5.

3.3.2.2 HELIX STABILISATION

From comparisons of the secondary structures of pig and *Tp. acidophilum* citrate synthase, helix stabilising properties in the archaeal protein have emerged. These include the presence of helix capping residues at the termini of the helices and also residues of high helical propensity. From the suggested secondary structure, it would appear that the archaeal enzyme has more capping residues and helix-capping box motifs at the termini of its helices than the pig protein. These residues

(a)



(b)

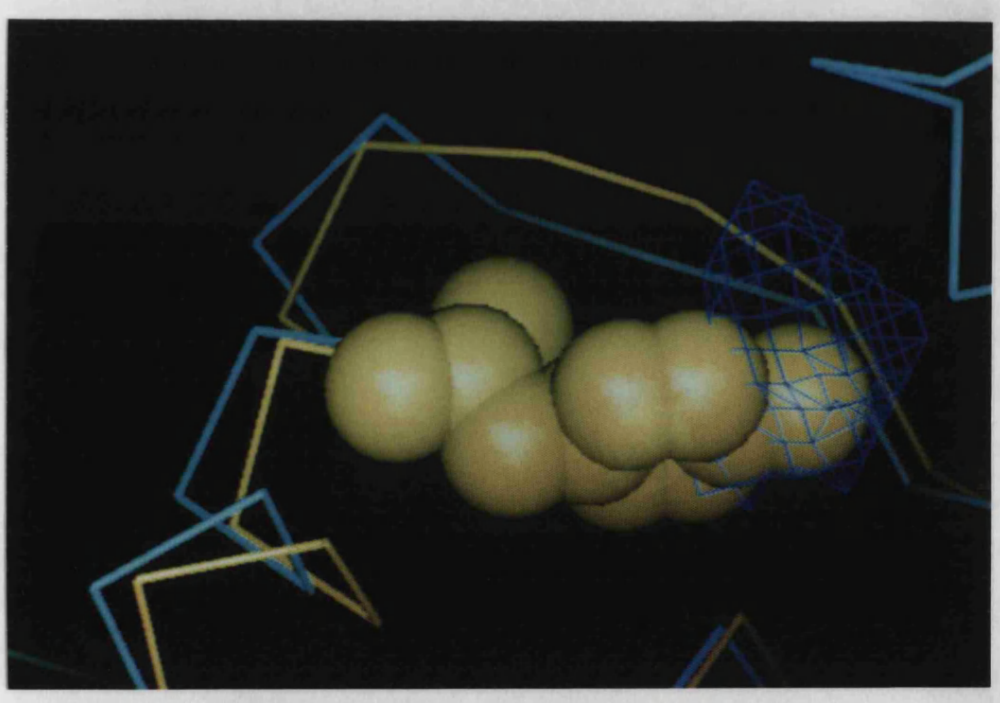


Fig. 3.9 One of the internal cavities (dark blue hashed area) of pig citrate synthase (blue α-carbon trace) is flanked by residue Ser 412 (blue, space-filled). In *Tp. acidophilum* citrate synthase (yellow α-carbon trace), the Ser is exchanged with residue Tyr 356 (yellow, space-filled), filling the cavity.

serve to stabilise the secondary structure through additional hydrogen bonding within the helix and counteraction of the helical dipole [Serrano *et al.*, 1992(c); Harper & Rose, 1993].

Helix G, at the subunit interface of the dimer of citrate synthase, may be another region that plays a significant role in conferring thermostability upon the protein. Half of Helix G in *Tp. acidophilum* citrate synthase contains Ala residues, as opposed to only 20% of the same helix in the pig enzyme. This feature may serve to stabilise helix G since Ala residues are the amino acid with the highest helical propensity [O'Neil & Degrado, 1990; Horovitz *et al.*, 1992].

3.3.2.3 AMINO ACID EXCHANGES

The multiple alignment of citrate synthase sequences (Fig. 3.1) highlighted four residues that are conserved in all the sequences except that from *Tp. acidophilum* (Table 3.2). The 3D positions of these residues have been determined. Two of the amino acids, Val 96 and Met 106, are opposite and antiparallel to the same residues in the other subunit at the subunit interface; although, due to their relative positions, they do not appear to interact with each other and thus they may be involved in alternative stabilising interactions. A third residue, Ser 132, is in the middle of helix I which is kinked in pig citrate synthase due to Pro 183, a conserved amino acid. Even though the kink is maintained in Helix I of the thermophilic structure, the absence of Pro may be significant, since Pro is the amino acid with the lowest helical propensity [O'Neil & Degrado, 1990; Horovitz *et al.*, 1992], and thus may aid destabilisation of the secondary structure in pig citrate synthase. This exchange of a Ser with a Pro in the thermophilic and mesophilic structures, respectively, has also been reported within a bent helix of phosphoglycerate kinase [Davies *et al.*, 1993].

The matrix of amino acid exchanges (Table 3.3) illustrates a number of features in *Tp. acidophilum* citrate synthase that could play a role in conferring thermostability. For example, there is a higher proportion of Ala residues that may help to stabilise the secondary structure of the protein (see section 1.3). This point is significant

when compared to pig citrate synthase where there are 25% fewer Ala amino acids. The fact that amino acid types Ser and Asn in the pig enzyme exchange with the Ala residues of the archaeal enzyme at a high frequency is in agreement with the protein stabilising traffic rules. Other features highlighted include the larger aromatic content of the archaeal protein, in the form of Phe, and also the increase in branched aliphatic residues, namely Ile. Both of these amino acid types are capable of raising the internal hydrophobic packing through additional interactions (discussed in section 3.3.2.4).

In the archaeal citrate synthase there is a significant increase in the number of Lys amino acids accompanied by a decrease in Arg residues. This feature is not in agreement with the earlier observations of Menendez-Arias & Argos [1989], who suggest, from comparative analyses of a range of thermostable and mesophilic proteins, that thermostability can be improved with a higher Arg/(Arg + Lys) ratio. Thus the feature reflects how the mechanisms for achieving stability are specific to individual proteins and a method that increases thermostability in one protein may not necessarily have a stabilising effect upon a different protein.

Other patterns gathered from the matrix include the decrease in Leu, the most prominent amino acid type in pig citrate synthase, together with a decrease in Asp, Gln and Asn, and the absence of all the Cys residues. The latter features are in agreement with the protein-stabilising traffic rules and appear significant since these residue types are irreversibly modified at elevated temperatures or can cause the cleavage of the polypeptide chain [Geiger & Clarke, 1987; Zale & Klibanov, 1986; Amaki *et al.*, 1994] .

3.3.2.4 ADDITIONAL INTERACTIONS

There is a marked increase in the number of aromatic interactions within the small domain of the archaeal citrate synthase when compared with the mesophilic enzyme; this has been highlighted in thermostable proteins previously [Burley & Petsko, 1985; Ishikawa *et al.*, 1993(a)].

Extra salt bridges have been correlated with an increase in the stability of proteins [Ishikawa *et al.*, 1993(a); Kelly *et al.*, 1993]. The contribution by ion pairs to the thermostability of *Tp. acidophilum* citrate synthase is not clear; The number of salt bridges in the archaeal enzyme both increases and decreases when using a 6Å and a 4Å cut-off distance between the charged groups, respectively [Russell *et al.*, in press].

3.3.2.5 HYPERTHERMOSTABILITY IN *P. FURIOSUS* CITRATE SYNTHASE

The crystal structure to 3.5Å of citrate synthase from the hyperthermophilic Archaeon *P. furiosus* has just been determined [Russell *et al.*, in press]. The structural comparisons between citrate synthases from pig and the moderate thermophile *Tp. acidophilum*, described above, can now be extended to include features that may confer hyperthermostability upon *P. furiosus* citrate synthase. Overall, the trends identified from mesophile to thermophile have been extended even further in the hyperthermophile. The *P. furiosus* enzyme appears to be the most compact structure of the three citrate synthases, with further truncation of its loop regions, elimination of all potential water-bearing cavities and an increase in the number at solvent-inaccessible atoms. In addition, the trends concerning the amino acid contents of the enzymes are extended, with a further reduction in the thermolabile residues Asn, Gln and Asp, and decrease in the Arg/Arg+Lys ratio. However, the increase and decrease in Ala and Leu residues respectively, is unique to *Tp. acidophilum* citrate synthase. The *P. furiosus* enzyme also resembles that from *Tp. acidophilum* with the absence of the N-terminal 36 residues and helix H.

Maintenance of dimeric integrity in the hyperthermostable citrate synthase appears to be have been achieved through two mechanisms: The degree of complementarity between the monomers at the interface increases with the growth temperature of the host organism. In addition there appear to be extra salt bridges at the subunit interface, particularly in the two central helical pairs, GG' and MM'. This shift to hydrophilicity in helix G of *P. furiosus* citrate synthase contradicts the significant hydrophobic content of the same helix in the moderately thermostable

protein. This feature may reflect the relative contributions of enthalpy and entropy to protein folding at different temperatures [Dill, 1990] and thus implies that the moderate thermophile and hyperthermophile have employed different mechanisms to achieve dimeric integrity.

Conducting comparative analyses of proteins from mesophiles and thermophilic Archaea with a view to characterising protein thermostability in these latter organisms is still in its infancy. Comparisons of proteins from both mesophilic and thermophilic Archaea have not even been reported as yet. It is highly likely that some of the features responsible for thermostability in archaeal proteins will agree with the general protein-stabilising traffic rules suggested (such as the increased Ala content of α-helices), others will be in disagreement (for example, the Arg/Arg+Lys ratio) and others may not yet have been revealed as novel features particular to the Archaea.

CHAPTER 4

INVESTIGATION OF INTERACTIONS AT THE SUBUNIT INTERFACE

4.1 INTRODUCTION

Citrate synthases from Eucaryotes, Gram positive Bacteria and the Archaea are all dimeric molecules, consisting of two identical subunits, both of which are essential for catalytic activity. It is thought that upon thermal denaturation of the dimeric pig enzyme, the monomers dissociate prior to their unfolding [McEvily & Harrison, 1986], and thus interactions at the subunit interface may be crucial for the integrity of dimeric citrate synthases from thermophilic organisms. As described in section 3.3.2, features have been identified that could play a role in stabilising citrate synthase from *Tp. acidophilum*. One such feature involves the nature and interactions of helices at the subunit interface of the protein. There is an increase in the Ala content of helix G in the archaeal enzyme, when compared with the equivalent helix in the pig enzyme. This secondary structure is part of one of four helical pairs at the subunit interface and thus the hydrophobic nature of the interactions at the interface has also been increased.

Mutational analysis and amino acid comparisons have shown that Ala is the residue with the highest helical propensity [O'Neil & Degrado, 1990; Horovitz *et al.*, 1992], and that its presence in α-helices, particularly at the subunit interface, is pronounced in thermostable proteins [Menendez-Arias & Argos, 1989]. Moreover, the presence of Ala residues can increase the thermostability of mesophilic proteins [Kotik & Zuber, 1993], and the most prolific exchange of residues in α-helices from mesophilic to thermophilic protein structures is from serine to alanine [Menendez-Arias & Argos, 1989]. All of the above features are connected with the increase in stability of the α-helix structure, due to the restriction in the conformational freedom

of alanine residues in the unfolded state and thus a smaller loss in entropy upon helix formation [O'Neil & Degrado, 1990]. In addition, studies mutating residues at the subunit interface of oligomeric proteins have shown that hydrophobic interactions at the interface are important for the stability of the thermophilic protein [Kotik & Zuber, 1993; Kirino *et al.*, 1994].

The effect of an increase in the alanine content of helix G, and thus an increase in the hydrophobicity at the subunit interface, of pig citrate synthase, is investigated in the present chapter by site-directed mutagenesis.

4.2 RESULTS

4.2.1 AMPLIFYING THE TEMPLATE PIG CITRATE SYNTHASE GENE

A pig citrate synthase gene was required as a template for use in the site-directed mutagenesis experiments. The clone pCS4, kindly donated by C. Evans (University of Texas, USA), contained the gene for pig citrate synthase preceded by a mitochondrial signal peptide [Evans *et al.*, 1988]. The signal peptide was removed by amplifying the remaining required portion of the gene using the polymerase chain reaction (PCR). The oligonucleotide primers were synthesised in-house (2.2.12) and are shown in Fig. 4.1. The restriction endonuclease sites *Eco*RI and *Hind*III were engineered into the 5' and 3' ends of the gene, respectively, and a translation initiation codon was engineered immediately following the *Eco*RI site and prior to the codon for residue Ala 1 of pig citrate synthase. An additional restriction endonuclease site, *Nco*I, that included the initiation codon, was also formed.

Forward PCR primer:

*Nco*I

5' (65) TT GCT GCC CG***G*** ***AAT*** ***T***CC ATG GCT TCT TCC (96) 3'

*Eco*RI start A1 S2 S3

Reverse PCR primer:

5' (1426) T TCT TCC CAG TCT G***AA*** ***GCT*** TCA CCC TCA CTT (1396) 3'

*Hind*III stop stop

Fig. 4.1 The forward and reverse oligonucleotide primers used in the PCR to remove the mitochondrial signal sequence from the pig citrate synthase gene. The primers read 5' to 3'. The base numbers are shown in brackets, the mutated residues are shown in bold and the endonuclease restriction sites are shown in italics (*Eco*RI and *Hind*III) or underlined (*Nco*I). The translation initiation codon (ATG) and termination codons (TCA), as well as the residues encoded, are also indicated.

The PCR was carried out as follows: 1, 10 and 100 ng template DNA (clone pCS4) were separately mixed with 1 μM of each primer, 50 μM of each dinucleotide triphosphate, 0.01% (w/v) BSA and PCR buffer (10 mM Tris.HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$). Control reactions were also carried out, omitting in turn the primer (10 ng template was used in the reaction) and the template DNA. The samples were incubated at 94°C for 5 min and then 1 U Taq DNA polymerase was added after which 25 cycles of PCR were carried out involving melting at 94°C for 90 s, annealing at 55°C for 105 s and extension at 72°C for 75 s. The success of the PCR was assessed on a 1 % (w/v) agarose gel (section 2.2.5) (Fig. 4.2), and then the amplified DNA was purified from a 1 % (w/v) LMP agarose gel (section 2.2.6). The gene was ligated into the vector M13mp19 at the *Eco*RI and *Hind*III restriction endonuclease sites and used to transfect *E. coli* strain DH5α, from which single stranded DNA was produced (protocols are described in sections 2.2.1 to 2.2.11). The entire pig citrate synthase gene was sequenced manually (section 2.2.13) to check for any random mutations and to make sure that the codon after the translation initiation signal coded for the first residue of the active enzyme. A single unintended mutation was found at base number 667, where there was a

G nucleotide in place of an A nucleotide; fortunately, the base change did not affect the amino acid encoded (Glu 196).

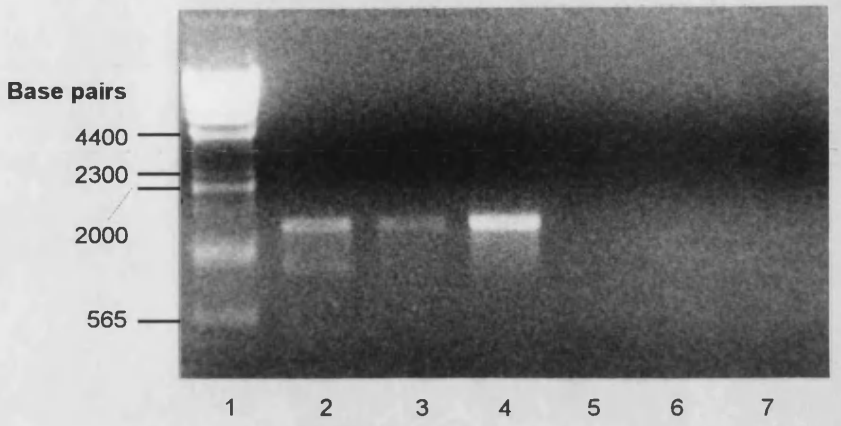


Fig. 4.2 **PCR to remove the mitochondrial signal sequence from pig citrate synthase.**
Lane 1: *Hind*III DNA size markers
Lanes 2-4: PCR products, using 1, 10, and 100 ng template DNA , respectively
Lanes 5-7: Control reactions omitting the forward primer, the reverse primer, and template DNA, respectively

4.2.2 MODELLING THE SUBUNIT INTERFACE MUTANTS OF PIG CITRATE SYNTHASE.

In pig citrate synthase, residues Ser 139 and Thr 146 are on the subunit interface side of helix G; their structurally equivalent residues in the enzyme from *Tp. acidophilum* are two Ala amino acids at positions 97 and 104, respectively. A double mutant of pig citrate synthase (S139A/T146A) was homology modelled (section 2.1), substituting the residues Ser 139 and Thr 146 with Ala residues. The open form of pig citrate synthase (1CTS.PDB) was used as the template. A superimposition of the α-carbon chains of the mutant structure on that of the wild-type structure showed that there was no change in the 3D positions of all the

Cα atoms. Ramachandran plots were identical between the two enzyme structures, indicating that the substitution of Ala amino acids for Ser 139 and Thr 146 had not introduced conformational strain in any of the residues throughout the molecule. There is a total of six hydrogen bonds that involve residues Ser 139 and Thr 146 per monomer of pig citrate synthase. Four of these bonds are formed between the backbone atoms of the two residues to other residues within the monomer, and these bonds are maintained in the mutant structure. The remaining two hydrogen bonds are formed between the side chains of Ser 139 and Thr 146 to amino acids in the other subunit in the dimeric enzyme; these bonds are obviously absent in the interface mutant of pig citrate synthase.

4.2.3 SITE-DIRECTED MUTAGENESIS TO CREATE SUBUNIT INTERFACE MUTANTS

The two subunit interface mutants of pig citrate synthase, T146A and S139A/T146A, were created by site-directed mutagenesis using the pALTER™ system (section 2.2.14). The mutagenic oligonucleotide used (Fig. 4.3) contained the mutations for both substitutions, and so only this one oligonucleotide was required to mutate both the Ser and Thr into Ala amino acids. Mutants were screened by manual sequencing (section 2.2.13) (results are shown in Fig. 1 of Appendix). The single mutant was created rather fortuitously, since upon screening for double mutants a clone was identified that contained the Thr 149 to Ala, but not the Ser 139 to Ala, mutation. This single mutant must have arisen as the result of inefficient binding of the mutagenic oligonucleotide to the wild-type template gene.

5' (489) T CTA CAC CCC ATG **G**CT CAG CTC AGT GCA GCC ATT **G**CA GCC CTC AAC AG (536) 3'
L134 H135 P136 M137 **A139** Q140 L141 S142 A143 A144 I145 **A146** A147 L148 N149

Fig. 4.3 Mutagenic oligonucleotide used to substitute Ala for Ser 139 and Thr 146, in the site-directed mutagenesis of pig citrate synthase. The base numbers are shown in brackets, the mutated bases are shown in bold and the residues coded for are shown below their respective codons.

4.2.4 EXPRESSION AND PURIFICATION OF THE SUBUNIT INTERFACE MUTANTS.

The wild-type and the two mutant pig citrate synthase genes were ligated into the expression vector, pKK223-3 at the *Eco*RI and *Hind*III restriction endonuclease sites (protocols are described in sections 2.2.2 to 2.2.8). Vector pKK223-3 carries the tac promoter for transcription, and ribosome binding site and termination signals for translation. The multiple cloning site possesses a unique *Eco*RI endonuclease restriction site that, if followed immediately by the translational initiation codon of a gene, will result in maximum efficiency of translation. The vector also codes for the β-lactamase gene endowing the host bacterium with ampicillin resistance. The recombinant vector was then used to transform the citrate synthase-deficient *E. coli* strain MOB154 (section 2.2.9).

Prior to purification of the recombinant pig citrate synthases, a cell extract (section 2.3.1) of the citrate synthase-deficient *E. coli* strain MOB154 was loaded on to and eluted off the Matrex™ gel Red A affinity column following the purification protocol described in section 2.3.5. The eluted fractions were assayed for an active citrate synthase (section 2.3.4) and also electrophoresed on 10% SDS-PAGE gels (section 2.3.3). The results indicated that no *E. coli* citrate synthase eluted from the column under the conditions that would be employed for elution of pig citrate synthase.

One ml volumes of overnight cultures of the three transformed *E. coli* clones (wild-type and mutants) were used to inoculate 2 l volumes of LB medium supplemented with ampicillin (100 mg/ml) (section 2.2.1). The cultures were incubated overnight at 37°C and then a cell extract was prepared from each (section 2.3.1). The citrate synthase was purified from each cell extract using the Matrex™ gel Red A affinity column, as described in section 2.3.5. The purification works on the principle that the Procion Red HE-3B dye, bound to the agarose beads in the Matrex™ gel Red A, has a high affinity for NADP-linked proteins, and since coenzyme A shares structural similarities with NADP, then the beads will also bind coenzyme A-linked proteins such as citrate synthase. The enzyme is then eluted using the ligands, coenzyme A and oxaloacetate.

The purity of the proteins was assessed by electrophoresis on 10% SDS-PAGE gels, together with the cell extracts and the flow through and wash solutions from the affinity columns (section 2.3.3) (Fig. 4.4). The cell extracts and the pooled fractions off the columns were assayed for citrate synthase activity at 25°C (section 2.3.4), and the amount of protein in each was calculated (section 2.3.2). The data are summarised in Table 4.1. The recombinant citrate synthases constituted 31%, 34% and 19% of the total cell protein for the wild-type, mutant T146A and mutant S139A/T146A, respectively.

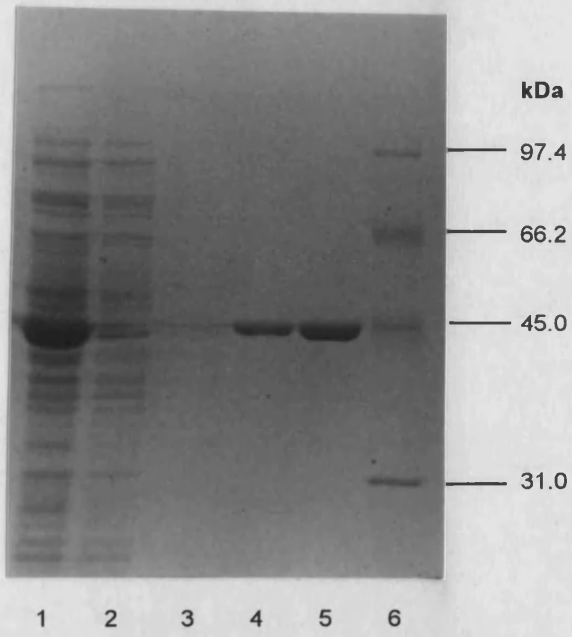


Fig. 4.4 **Purification of citrate synthase**

The results of the SDS-PAGE analysis of the purification of the wild-type and mutant pig citrate synthases were identical, and thus only the gel of the single mutant purification samples is shown.

Lane 1: Cell extract of *E. coli* MOB154 strain carrying the pKK223-3 vector-encoded mutant T146A gene (20 μg)

Lane 2: Flow through from the Matrex™ gel Red A affinity column whilst loading the cell extract

Lane 3: Wash from the above column

Lane 4: Fraction containing the highest citrate synthase activity upon elution from the affinity column (1 μg)

Lane 5: Pooled sample of purified mutant citrate synthase (1.5 μg)

Lane 6: Protein standards, with the molecular weights indicated

PROTEIN	STEP	TOTAL ENZYME ACTIVITY (μmol/min)	TOTAL PROTEIN (mg)	SPECIFIC ACTIVITY (μmol/min/mg)	YIELD (%)
Wild-type	Cell extract	8,107	246	33	100
	Purified	7,854	74	106	97
T146A	Cell extract	5,525	266	21	100
	Purified	5,105	84	61	92
S139A/T146A	Cell extract	4,886	500	10	100
	Purified	4,486	84	53	92

Table 4.1 Purification of wild-type and mutants T146A and S139A/T146A of pig citrate synthase on the affinity Matrex™ gel Red A column.

The coenzyme A and oxaloacetate were removed from the solutions of purified proteins using a G-25 Superose gel filtration column, as described in section 2.3.5. These protein solutions were used for further characterisation studies.

4.2.5 KINETIC CHARACTERISATION OF THE SUBUNIT INTERFACE MUTANTS

The kinetic parameters (true Km and Vmax values) were calculated for the wild-type pig citrate synthase and the two mutants of the enzyme, T146 and S139A/T146A. This was achieved by measuring the citrate synthase activity, at 37°C (section 2.3.4), in conditions that employed a range of concentrations of both substrates, acetyl coenzyme A and oxaloacetate, from approximately 0.15 Km to 13 Km. The protein concentration used in each assay was 37 ng/ml for wild-type pig citrate synthase, 28.5 ng/ml for mutant T146A, and 193 ng/ml for mutant S139A/T146A, and each assay was carried out in triplicate.

The dependency of enzymic rate on oxaloacetate concentration at various fixed concentrations of acetyl coenzyme A was determined. The experiment was repeated analysing the dependency on acetyl coenzyme A concentration at various fixed concentrations of oxaloacetate. The values of $Vmax^{app}$ and Km^{app} at the fixed

concentrations of acetyl coenzyme A and oxaloacetate were calculated, respectively, from direct linear plots. From secondary direct linear plots (Fig 4.5), the true values of Km and Vmax were determined (Table 4.2).

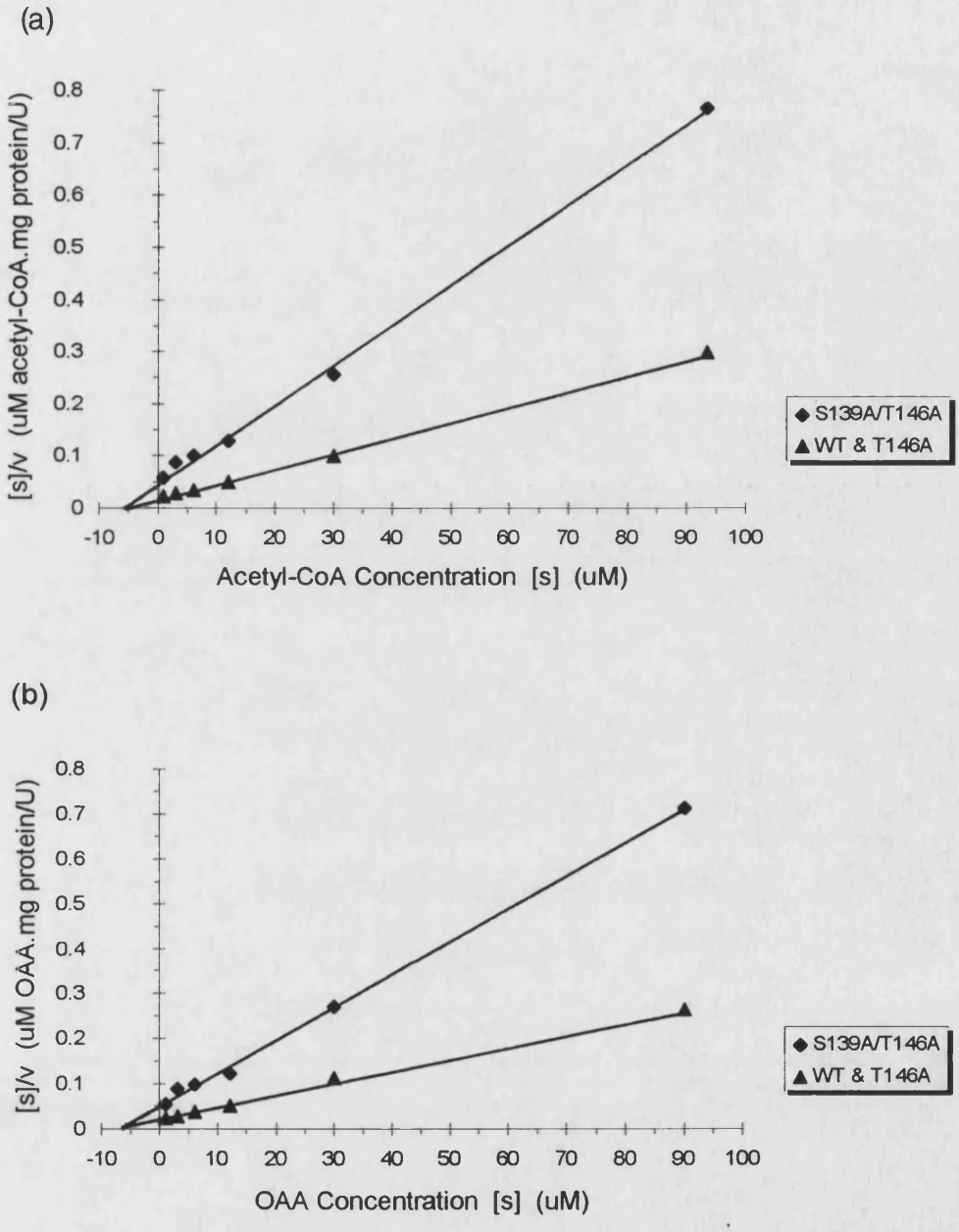


Fig. 4.5 Hanes Woolf plot of (S/v) versus (S) where S is the substrate concentration (μM) and v is the apparent Vmax (μmol/min/mg protein) for (a) acetyl coenzyme A, and (b) oxaloacetate, for the wild-type pig citrate synthase enzyme and the interface mutant S139A/T146A. The plots for mutant T146A are almost identical to that of the wild-type enzyme and so have not been shown. The true Km and Vmax parameters have been calculated from direct linear plots, but the Hanes-Wolf plot is shown for level of clarity.

ENZYME	TRUE Km (μM)		TRUE Vmax (μmol/min/mg)
	OAA	AcCoA	
Wild-type 68% confidence limits	7.0 (6.2-7.2)	7.1 (6.4-7.7)	370 (340-400)
T146A 68% confidence limits	4.8 (3.7-6.0)	5.5 (4.7-6.6)	360 (330-410)
S139A/T146A 68% confidence limits	7.5 (6.5-9.3)	7.5 (5.9-9.6)	140 (130-160)

Table 4.2 True Km and Vmax values for substrates oxaloacetate (OAA) and acetyl coenzyme A (AcCoA), for the wild-type pig citrate synthase and the two mutants, T146A and S139A/T146A, of the enzyme. The values were calculated from direct linear plots and the 68% confidence limit values are also shown.

As can be seen from Table 4.2, the true Km values for both substrates are very similar in all three enzymes, ranging from 4.8-7.5 μM. The Vmax values in the double mutant (140 U/mg) are less than half that of those in the wild-type citrate synthase and the single mutant (approximately 370 U/mg).

4.2.6 CHARACTERISATION OF THE THERMAL INACTIVATION OF THE INTERFACE MUTANTS

The thermal inactivation of a protein can be characterised using the Arrhenius equation shown below:

$$\frac{d\ln(k)}{dT} = \frac{E_a}{RT^2} \qquad \text{Eqtn. (1)}$$

Where,

k is the first order rate constant for inactivation,

T is temperature in Kelvin,

E_a is the activation energy,

R is the gas constant (8.314 J/mol/K).

The activation energy can be assumed to be independent of temperature and so integration of the Arrhenius equation gives:

$$k = A\, e^{-E_a/RT} \qquad \text{Eqtn (2)}$$

For a thermal inactivation reaction, the value of E_a is the energy of activation of the inactivation process, and can be determined from a plot of ln(k) against 1/T, known as the Arrhenius plot, the gradient of which is $-E_a/R$ (Eqtn 3).

$$\ln(k) = \ln(A) - \frac{E_a}{R}.\frac{1}{T} \qquad \text{Eqtn (3)}$$

The thermal inactivation of the interface mutants of pig citrate synthase was characterised and compared with that of wild-type citrate synthase. The proteins were incubated at a concentration of 10 μg/ml (0.102 μM), in 2TE8 buffer, at a range of temperatures from 31°C to 43°C. At noted time intervals, 50 μl volumes of the samples were placed on ice for 30 s from which a 20 μl aliquot was assayed for enzyme activity at 37°C in a 1 ml assay volume (section 2.3.4). Plots of ln(% activity remaining) against time were drawn (Fig. 4.6). The half-life, at 37°C, for mutant T146A was 29 min and for mutant S139A/T146A it was reduced to 18 min. When wild-type citrate synthase was incubated at 37°C there was 55% activity remaining after 60 min. The rate constants for inactivation at each temperature were calculated and used to draw Arrhenius plots (Fig. 4.7); however, as can be seen from Fig. 4.6(a), it was difficult to determine values for the rate constants for wild-type citrate synthase since the inactivation was not a standard single first-order process. The non-linearity of the inactivation data for the wild-type enzyme is investigated in section 4.2.7. For the mutant proteins, the values for the energy of activation of the inactivation process (E_a) were calculated to be 200 kJ/mol for T146A, and 189 kJ/mol for S139A/T146A.

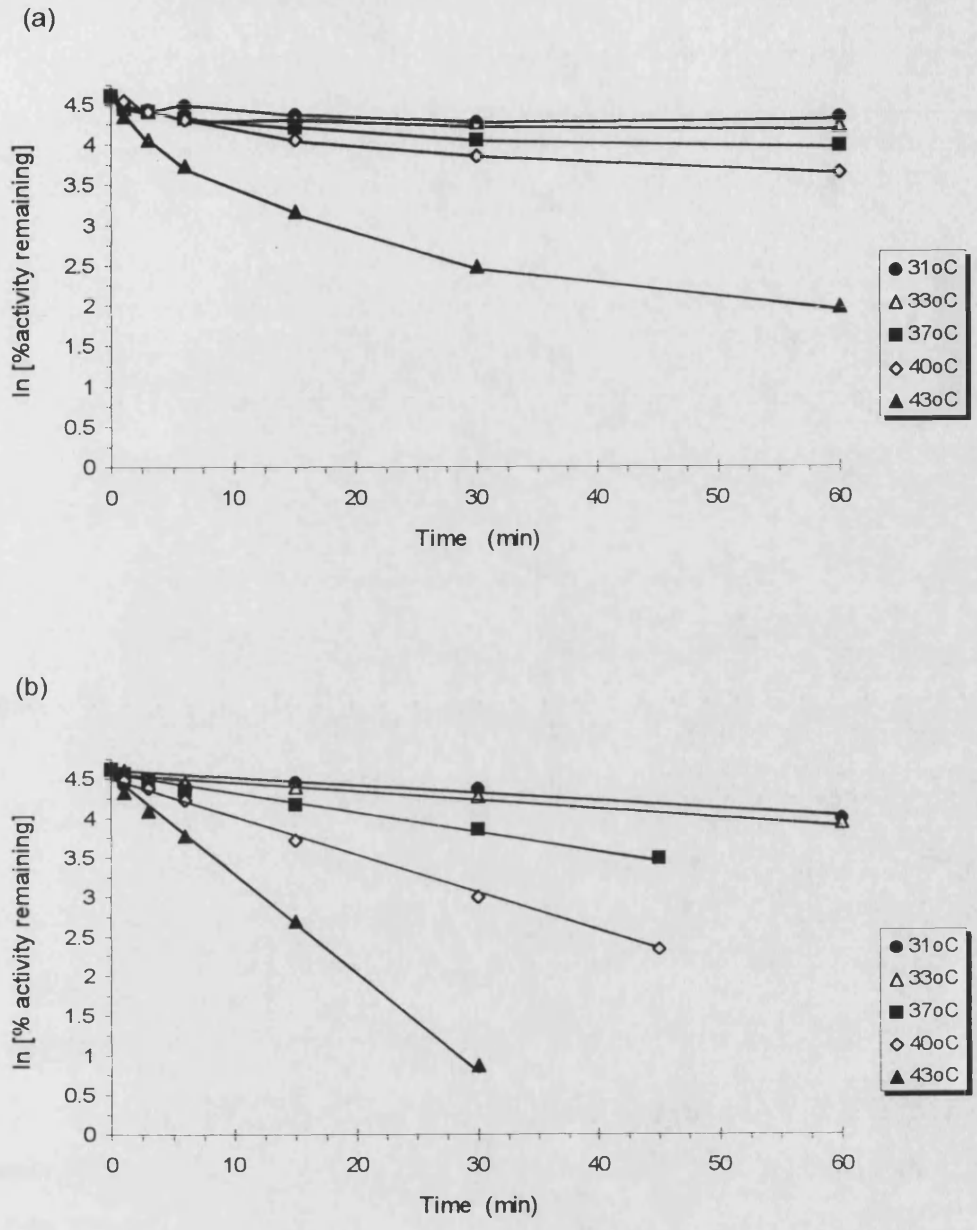


Fig. 4.6 **Thermal inactivation of citrate synthases at a range of temperatures as a function of time.** The gradient of each line equates to the rate constant for inactivation (k).

(a) Wild-type pig citrate synthase

(b) Single mutant, T146A, of citrate synthase

(c)

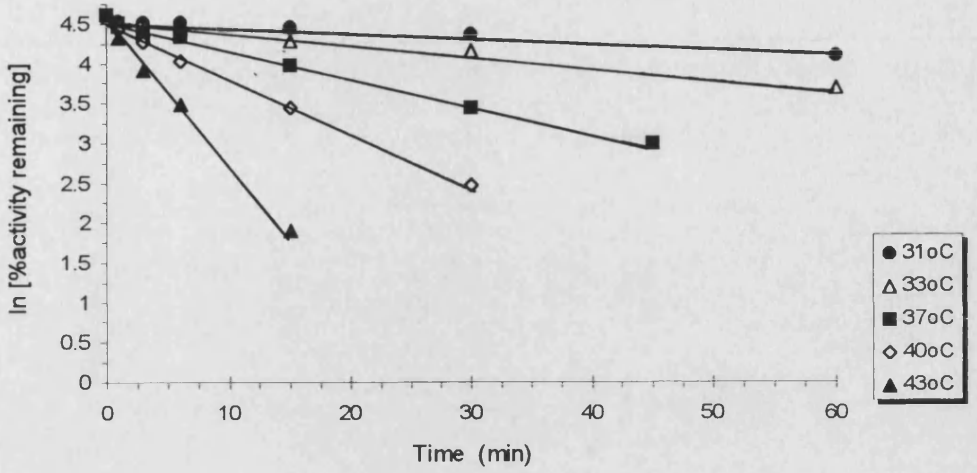
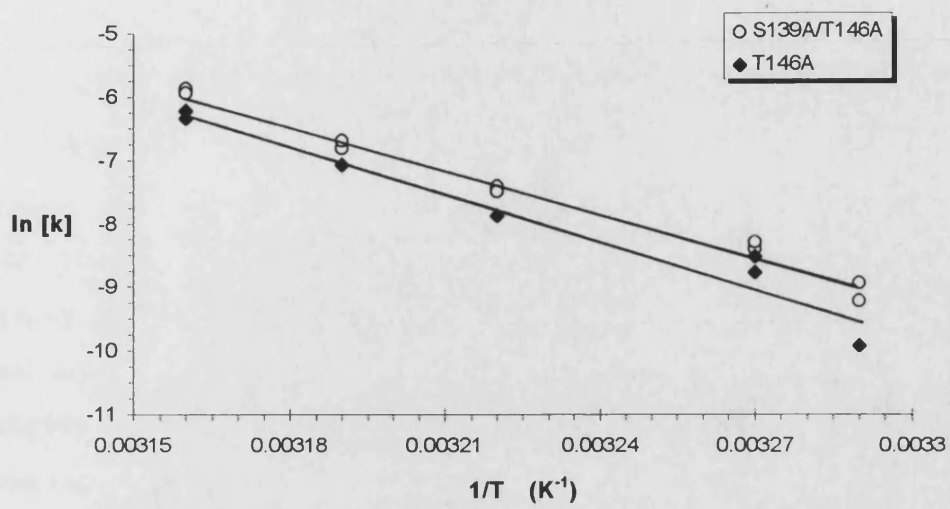


Fig. 4.6**(c)** **Thermal inactivation of citrate synthases at a range of temperatures as a function of time.** The gradient of each line equates to the rate constant for inactivation (k). Double mutant, S139A/T146A, of citrate synthase.

4.2.7 INVESTIGATING THE EFFECT OF PROTEIN CONCENTRATION UPON THERMAL DENATURATION OF CITRATE SYNTHASE

Research on the subunit equilibrium of pig citrate synthase [McEvily & Harrison, 1986] has suggested that the denaturation of the enzyme is dependent on protein concentration. Thus at concentrations in excess of 5 μM, there was no loss of citrate synthase activity after 6 h at 35°C; however, as the concentration was lowered the rate of inactivation increased, and at 5 nM there was almost complete loss of activity after 3 h. The intermediate concentrations exhibited a biphasic loss of activity, and a mathematical model was proposed in which the increase in the concentration of denatured monomer was dependent on two consecutive first-order processes: the dissociation of dimers (rate constant k_{-1}), and the denaturation of monomers (rate constant k_2), where k_2 was much greater than k_{-1}.

(a)



(b)

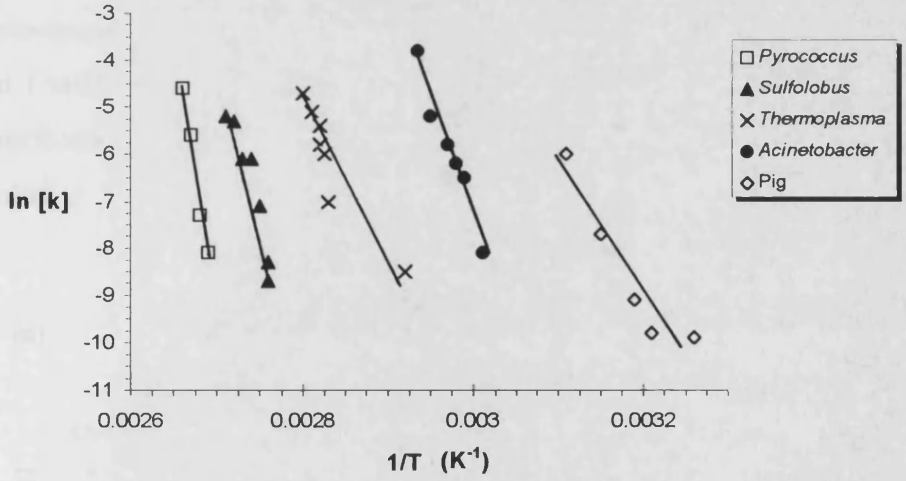


Fig. 4.7 **Arrhenius plots:**

(a) the single mutant, T146A, and the double mutant, S139A/T146A, of citrate synthase.

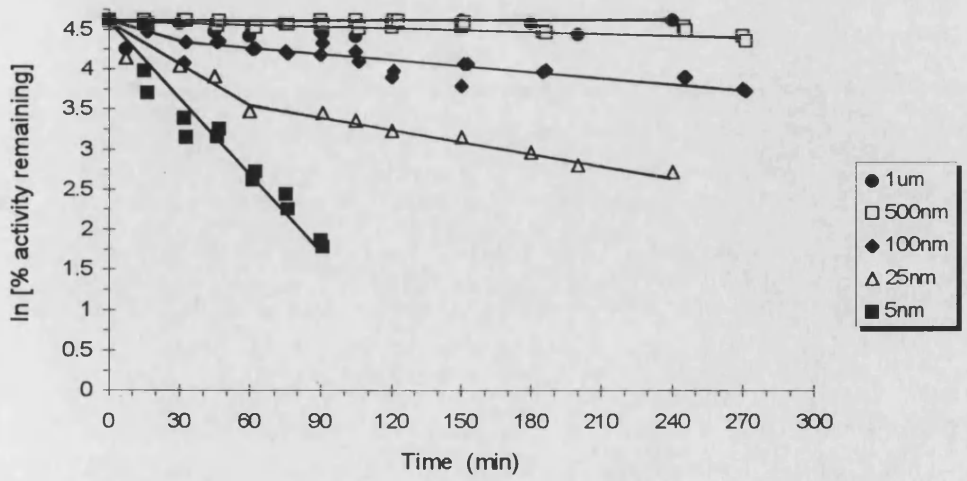
(b) a range of citrate synthases, emphasising the relative stabilities of the enzymes from the Archaea (Muir *et al.*, 1995).

The activation energy of the inactivation process (E_a) can be calculated from the gradient of the plot.

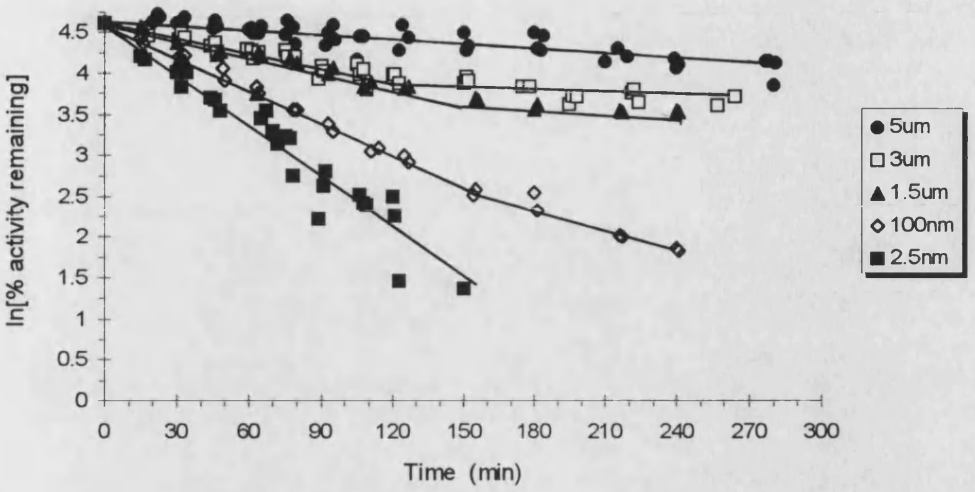
The thermal denaturation of pig citrate synthase in the previous section was carried out at a protein concentration of 100 nM. The plots of the denaturation appeared to be of a single first-order process for the two interface mutants of the enzyme (Fig. 4.6(b) and (c)); however, the thermal denaturation of wild-type pig citrate synthase was of a non-linear nature (Fig. 4.6(a)). The non-linearity of the latter plots may be the same as the biphasic loss of activity seen with pig citrate synthase by McEvily & Harrison [1986] since the loss of activity at the protein concentration employed in the thermal denaturation studies in section 4.2.6 also showed biphasic loss in the reported work, although the reaction conditions used were slightly different. The dependency of denaturation upon protein concentration was investigated in both wild-type pig citrate synthase, and in the interface mutants, in order to highlight the obvious differences in thermal denaturation between the three proteins.

The citrate synthases were incubated in 2TE8 at 35°C and, at noted time intervals, aliquots of the samples were assayed for activity at 35°C (section 2.3.4). The range of concentrations used for wild-type citrate synthase was 5 nM-1 μM, for the single mutant T146A, 2.5 nM-5 μM, and for the double mutant S139A/T146A, 25 nM-2 μM. The inactivation plots for all three proteins are shown in Fig. 4.8. The data are also presented in Table 4.3 and Fig. 4.9.

(a)



(b)



(c)

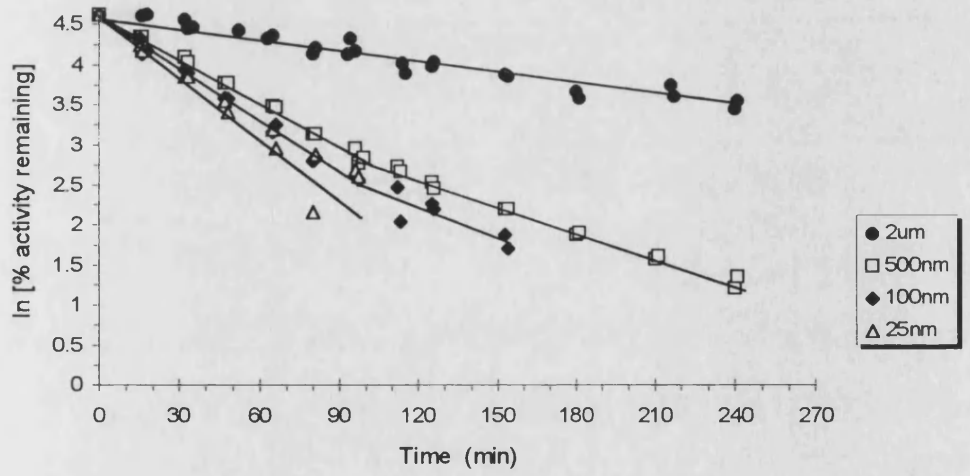


Fig. 4.8 **The effects of protein concentration upon inactivation of citrate synthases at 35°C.**

(a) Wild-type pig citrate synthase

(b) Mutant T146A

(c) Mutant S139A/T146A

WILD-TYPE		MUTANT T146A		MUTANT S139A/T146A	
Conc. (μM)	Half-life (min)	Conc. (μM)	Half-life (min)	Conc. (μM)	Half-life (min)
1	(100%)	5	420	2	158
0.5	1243	3	120	0.5	40
0.1	201	1.5	111	0.1	33
0.025	38	0.1	53	0.025	32
0.005	25	0.0025	34		

Table 4.3 Half-life of pig citrate synthases, at 35°C, at varying protein concentrations. If necessary the plots were extrapolated in order to calculate the half-lives. For 1 μM wild-type protein concentration, the activity remaining after 4 h was 100%, and thus the half-life was incalculable.

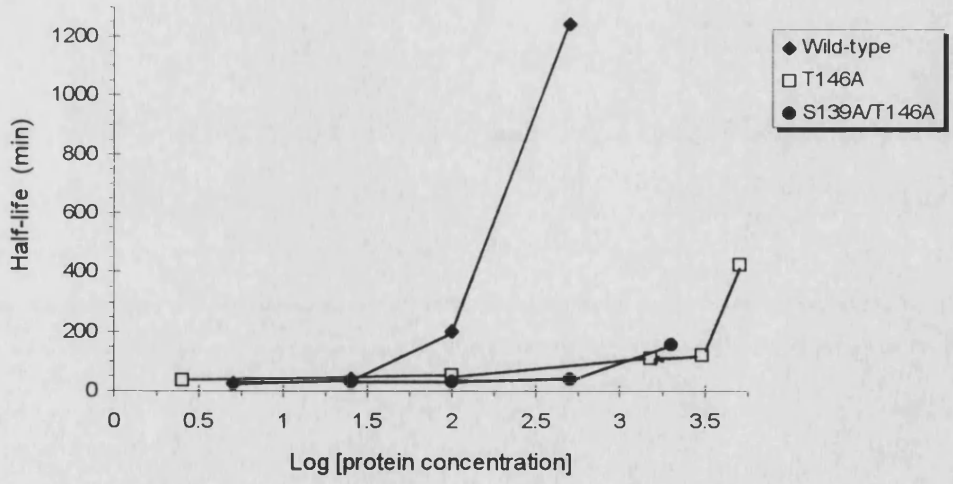


Fig. 4.9 Half-life of pig citrate synthases, at 35°C, at varying protein concentrations

4.2.8 DETERMINING THE MELTING TEMPERATURE OF THE INTERFACE MUTANTS

The melting temperature (T_m) of a protein is the point at which the structure unfolds in solution and is an indication of the thermal stability of that protein. Differential scanning calorimetry (DSC) is a technique that can be used to determine the T_m of a protein in solution. The diagram below (Fig. 4.10) illustrates the DSC results obtained with a theoretical protein.

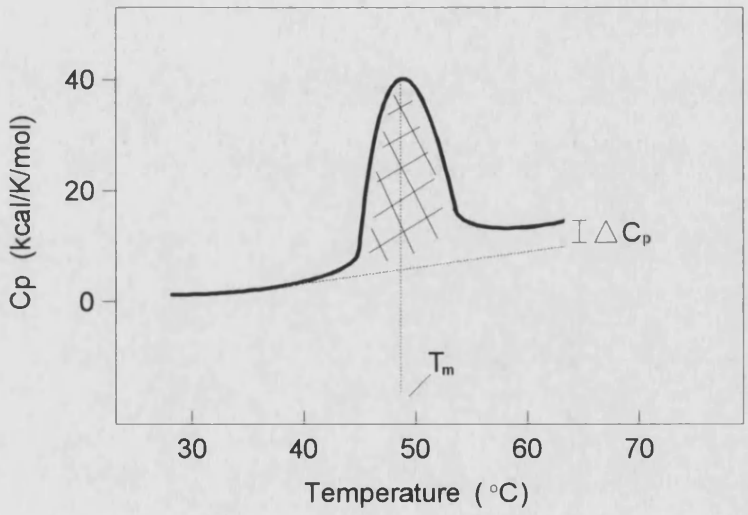


Fig. 4.10 Differential scanning calorimetry results obtained with a theoretical protein. The melting temperature (T_m) and the change in the partial heat capacity (ΔC_p) are indicated. The shaded area relates to the enthalpy change upon unfolding (ΔH)

The sample is heated to a defined temperature and the amount of energy required to reach this point, termed as the heat capacity (C_p), is noted. The temperature is then raised and the C_p noted at each increment. The contribution of the protein to the C_p, known as the partial C_p, is determined by subtracting the corresponding measurements of the aqueous solvent. The partial C_p is a combination of that of the protein and also any effects that the protein has on the surrounding solvent.

Initially, the increase in temperature has only a slight effect upon the partial C_p of the folded protein. Unfolding of the protein is then apparent as a peak in the heat capacity curve, indicating a large absorption of energy. The area underneath the transition peak gives the enthalpy change (ΔH) upon unfolding. Above the T_m, the partial C_p becomes more constant again, but the unfolded state retains a greater heat capacity than the folded state. The difference between the two values has been shown to be approximately proportional to the buried non-polar surface area that is assumed to be exposed upon unfolding [Privalov & Makhatadze, 1990]; this is expected since the value of the partial C_p of a protein is dominated by the level of non-polar surface area exposed to water [Makhatadze & Privalov, 1990].

Differential scanning calorimetric studies (carried out by Dr A. Cooper, University of Glasgow) were used to determine the T_m of the wild-type citrate synthase and the two mutants, T146A and S139A/T146A, of the enzyme. The protein solutions, all in 2TE8 buffer, were analysed at constant pressure and concentrations of 2.64 mg/ml for the wild-type enzyme, 2.9 mg/ml for mutant T146A, and 2.75 mg/ml for mutant S139A/T146A (measured spectrophotometrically at 280 nm using the absorption coefficient, $\varepsilon^{0.1\%}_{1cm}$=1.78). The temperature was raised at a rate of 60°C/h. The results of the DSC are shown in Fig. 4.11. For the wild-type citrate synthase, the T_m was found to be 50°C, whereas in the two mutants this value was slightly lower at 47.5°C. In addition, there seemed to be a second transition at 57°C for the wild-type enzyme and the double mutant, S139A/T146A. In each case, the partial C_p of the unfolded protein appeared to be the same as the value for the folded protein.

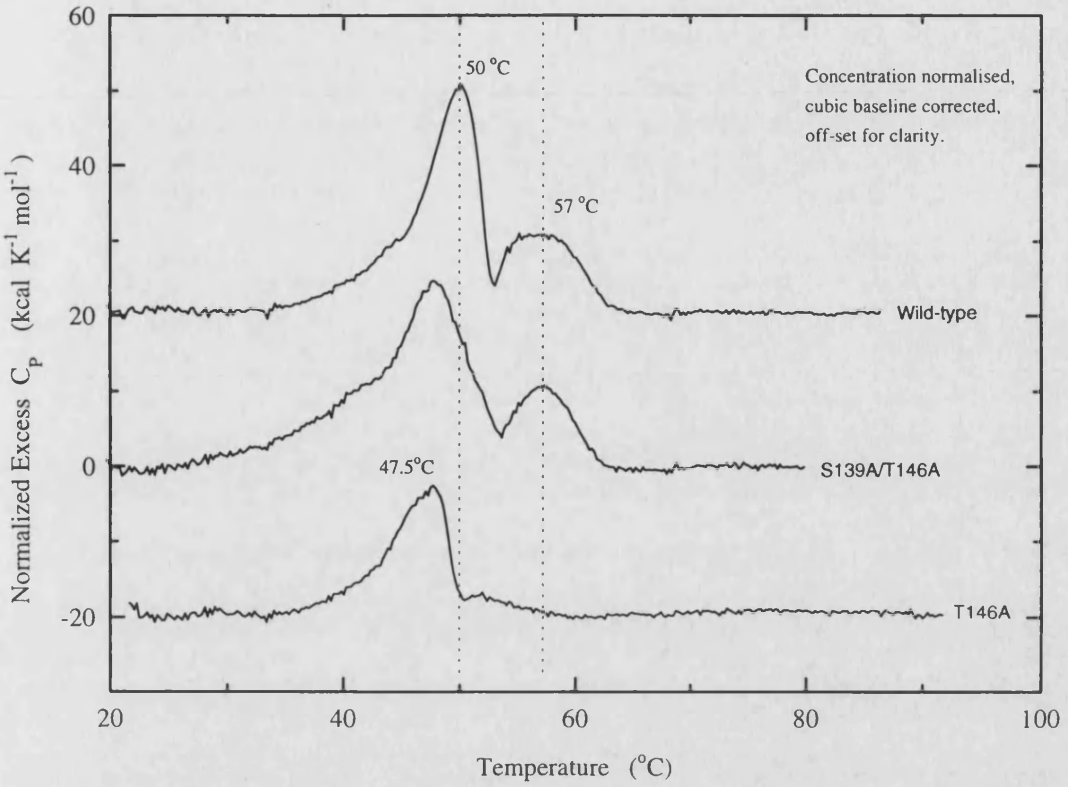


Fig. 4.11 Differential scanning calorimetry data measuring the heat capacity (C_p) of wild-type pig citrate synthase and the two interface mutants of the protein.

4.2.9 STRUCTURAL ASSESSMENT OF THE INTERFACE MUTANTS

Circular dichroism spectroscopy (CD) can be used to assess the tertiary and secondary structure of a protein. This form of spectroscopy involves the passage of right- and left-handed circularly polarised light through a solution of optically active asymmetric molecules. The two forms of the light will be absorbed to unequal degrees and it is this differentiation in absorption coefficients that is measured. Proteins are optically-active due to the presence of L-amino acids and also to asymmetry of the overall folded conformation. There are two spectral regions for a protein. The far-UV, or amide region (170-250 nm), detects the contribution from peptide bonds. Here, the secondary structural aspects of a protein are defined, α-helices in particular have a strong and characteristic CD spectrum, whereas other

secondary structural elements are less well defined. The second region is the near-UV (250-300 nm) where the contribution from aromatic amino acids is detected. The spectrum in this latter region is very complex since the signal from each aromatic residue depends on the immediate structural and electronic environment around that residue, thus the transition of an individual residue cannot be defined. Instead, the near-UV region is very sensitive to the overall native state of the protein, highlighting changes to the native state as a function of the environment of the aromatic residues within the structure.

In order to determine whether the substitution of Ser 139 and Thr 146 residues for Ala amino acids had affected the tertiary and/or the secondary structure of pig citrate synthase it was necessary to carry out structural analyses on the mutant proteins and compare the data with the wild-type structure. Circular dichroism studies (carried out by N.C. Price, University of Stirling) were performed using a JASCO J-600 spectropolarimeter. The far-UV and near-UV CD spectra were recorded at 20°C for the wild type pig citrate synthase and the two interface mutants of the enzyme (Figs. 4.12 and 4.13, respectively). The working concentrations of the protein samples, all in 2TE8 buffer, were 0.151 mg/ml for the wild-type enzyme, 0.145 mg/ml for mutant T146A, and 0.152 mg/ml for mutant S139A/T146A.

The secondary structure content of the three proteins was estimated from the far-UV spectra, using data between 195-240 nm, with the CONTIN program [Provencher & Glöckner, 1991] (Table 4.4). This program makes use of a reference set of structurally-characterised proteins for the predictions. From the far-UV spectra (Fig. 4.12, 225 nm wavelength) and the data in Table 4.4, it would appear that the two citrate synthase mutants may have a lower helical content than the wild-type protein. As far as the near-UV spectra are concerned, the peak at 295 nm, corresponding to the environment of Trp residues, is slightly shifted for the two mutant structures compared with that for wild-type citrate synthase; this indicates that there are subtle differences in the tertiary structure of the mutants compared with the wild-type protein structure.

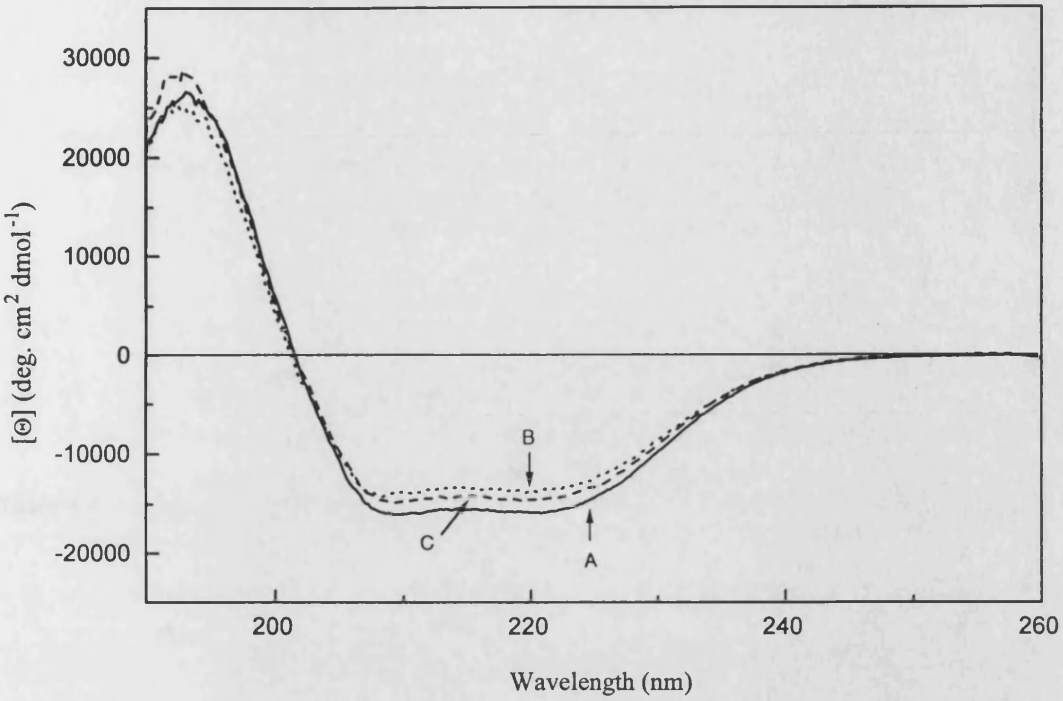


Fig. 4.12 Far-UV spectra recorded for wild-type pig citrate synthase (A), mutant T146A (B), and mutant S139A/T146A (C). The degree of α-helical content is indicated by the reading at 225 nm.

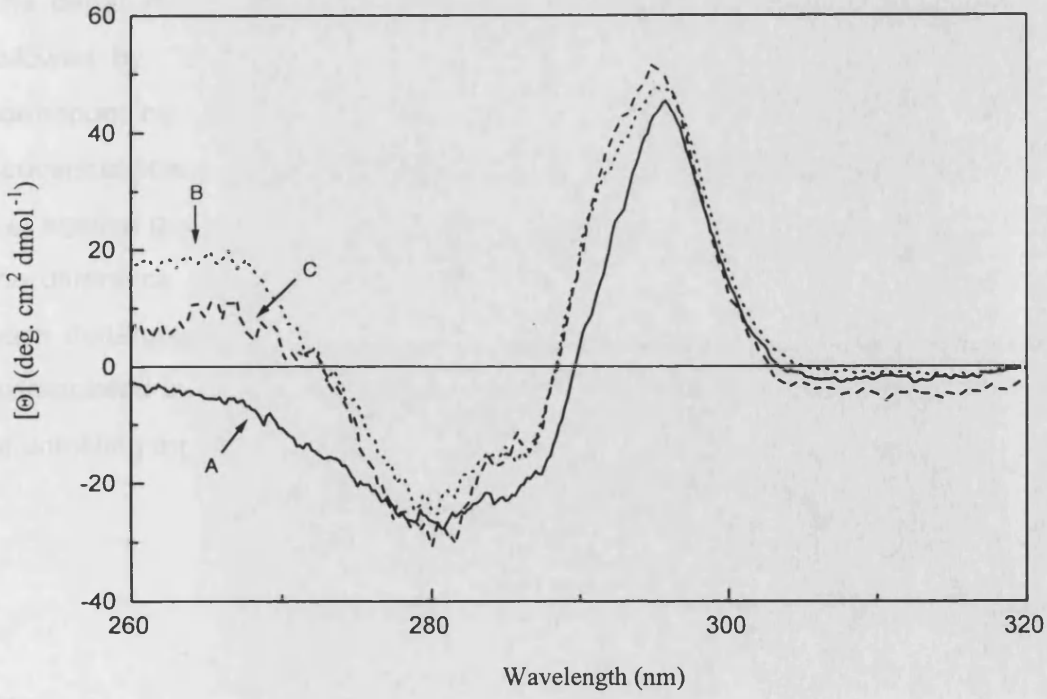


Fig. 4.13 Near-UV spectra recorded for wild-type pig citrate synthase, (A), mutant T146A (B), and mutant S139A/T146A (C). The peak at 295 nm relates to the environment of Trp residues.

PROTEIN	% α-HELIX	% β-SHEET	% OTHER
Wild-type	55 ± 1	20 ± 2	25 ± 3
T146A	42 ± 2	26 ± 3	32 ± 4
S139A/T146A	41 ± 2	18 ± 3	41 ± 4
Actual wild-type	72	3	25

Table 4.4 Secondary structure content of wild-type pig citrate synthase, and the interface mutants, T146A, and S139A/T146A, estimated from the far-UV CD spectra (Fig 4.12). The actual secondary structure content of the wild-type protein, as determined from the crystal structure [Remington *et al.*, 1982], is also given.

4.2.10 GUANIDINE DENATURATION OF THE CITRATE SYNTHASES

The denaturation of a protein structure by treatment with guanidine.HCl can be followed by CD (section 4.2.9). The change in CD spectrum, θ, at 225 nm, corresponding to the α-helical content, is monitored over a range of denaturant concentrations. With the use of an extensive and accurate set of data, the plot of θ_{225} against guanidine.HCl concentrations can be used to calculate a value for ΔG, the difference in free energy between the folded and unfolded conformations, at each denaturant concentration, using equation 4. The value of ΔG can then be extrapolated back to 0 M guanidine.HCl, i.e. water, in order to calculate the $\Delta G_{(H_2O)}$ of unfolding for the protein [Pace, 1990].

$$\Delta G = -RT \ln\left[\frac{F_d}{(1-F_d)}\right] \qquad \text{Eqtn (4)}$$

Where,

F_d is the fraction of protein denatured (the value of θ_{225} for the native conformation minus the observed value for θ_{225})

The denaturation by guanidine.HCl of pig citrate synthase and the two subunit interface mutants was followed by CD spectroscopy. The proteins, at the concentrations indicated in section 4.2.9, were incubated with guanidine.HCl, at concentrations ranging from 0-6 M, for 15 min at room temperature. The CD spectra were then recorded for each sample (Fig. 4.14), and the change at 225 nm was noted (Fig. 4.15). These data are not thorough enough to calculate the ΔG values at each denaturant concentration, since the dependence of θ_{225} on guanidine.HCl concentration cannot be defined accurately. From an analysis of Fig. 4.15, it appears that there may be very subtle differences between the effects of denaturant upon the structure of the wild-type and mutant proteins, although at this stage of the analysis any differences between the unfolding of the three protein structures are very small, if they exist at all, and a more thorough analysis is required over the whole range of guanidine.HCl concentrations before any sound conclusions can be made.

4.2.11 CRYSTALLISATION TRIALS OF THE INTERFACE MUTANTS.

Knowledge of the 3D structures of the interface mutants of citrate synthase would be very useful in comparative analyses with the X-ray structure of the wild type pig enzyme [Remington *et al.*, 1982]. Such an analysis could indicate if, and how, the substitutions of residues Ser 139 and Thr 146 for Ala amino acids had affected the structural conformation of the surrounding residues, the conformation of helix G, and also the interactions at the subunit interface. This information would then be used to shed light on the results of the kinetic, thermodynamic and spectral analyses carried out in previous sections of this chapter.

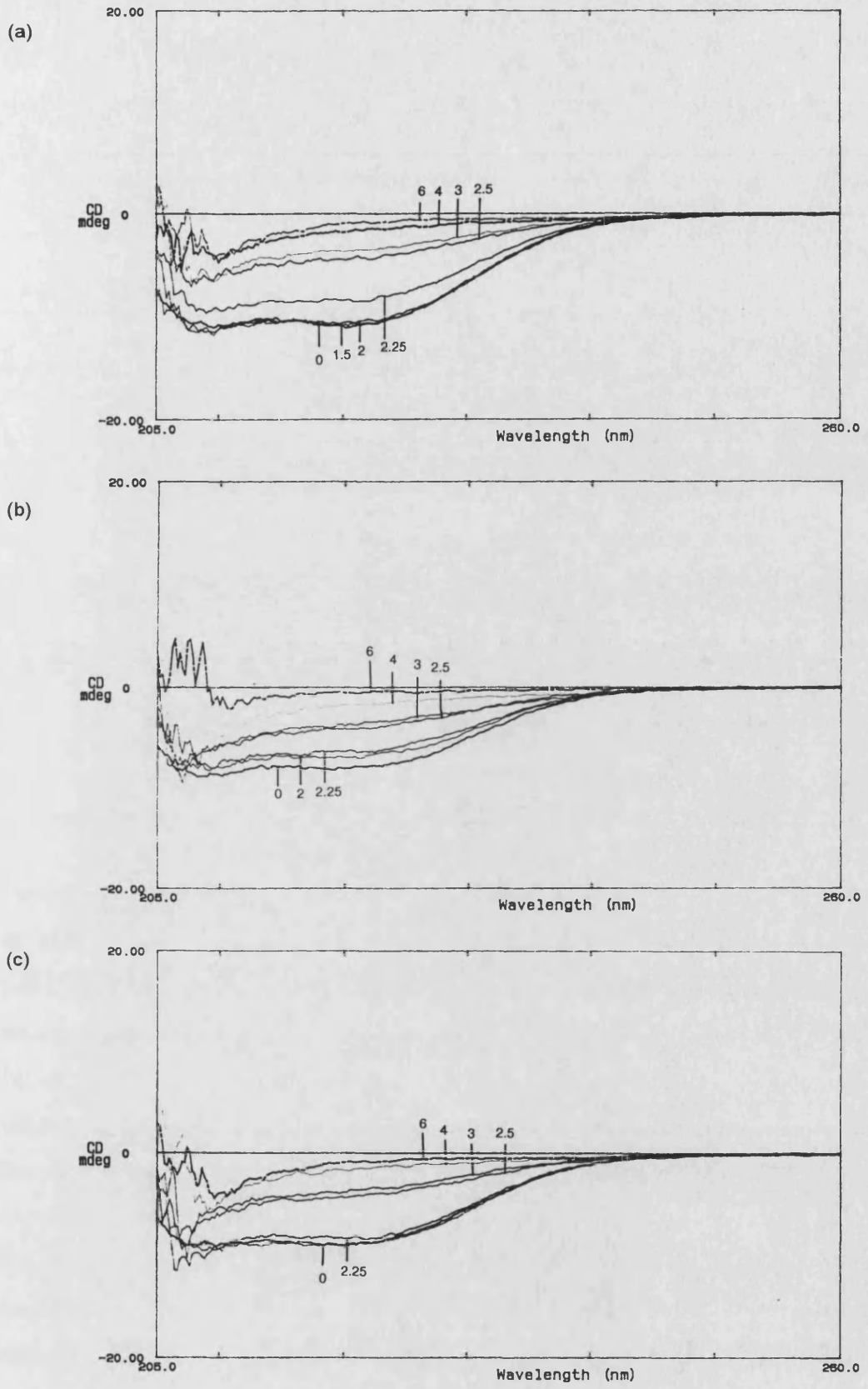


Fig. 4.14 Circular dichroism spectra following protein denaturation by guanidine.HCl (concentrations, in M, are indicated on each spectrum), recorded at 20°C, for (a) wild-type pig citrate synthase, (b) mutant T146A and (c) mutant S139A/T146A of the enzyme.

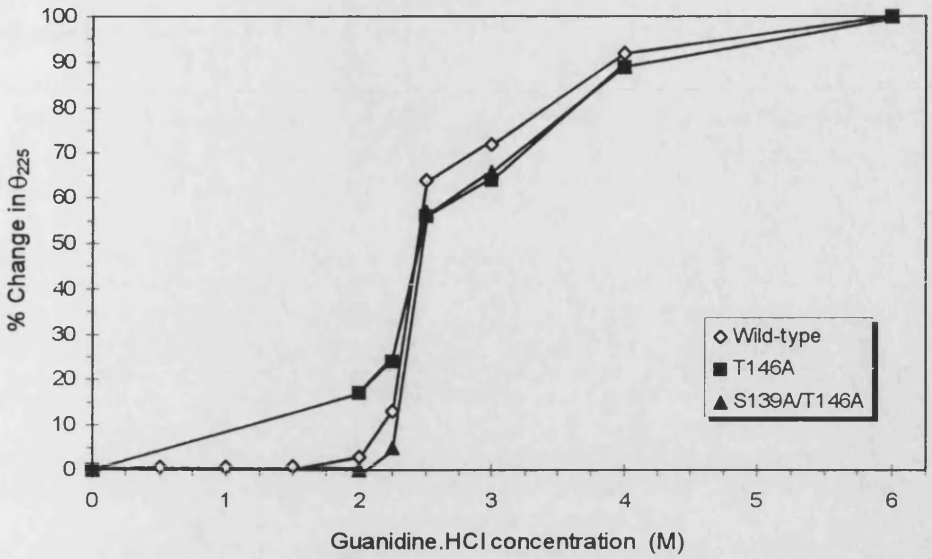


Fig. 4.15 The percentage change in spectral reading at 225 nm as a function of guanidine.HCl concentration for wild-type pig citrate synthase and the mutants T146A and S139A/T146A. The readings at 0 M and 6 M denaturant concentrations have been taken to represent 100% and 0% native conformation, respectively.

Protein crystallisation trials were set up for mutant T146A and mutant S139A/T146A of pig citrate synthase. Using Amicon Centriprep concentrators, a 3 mg sample of each of the two proteins, purified on the affinity column (section 4.2.4), was buffer-exchanged into crystallisation buffer (0.9 M sodium citrate, pH 6.0, 1 mM coenzyme A) and concentrated to 7 mg/ml [Remington *et al.,* 1982]. The hanging drop, vapour diffusion method was then employed in the crystallisation trials. The buffers used in the reservoir included 1.2 M citrate and also those of the Magic 50 Crystal Screen. After 7 days, a mass of tiny needle-like crystals were seen in the protein drop using Buffer 39 of the Screen, 0.1 M Na.HEPES, pH 7.5, 2 % (v/v) PEG400, 2 M ammonium sulphate. These crystals were too small to mount onto the X-ray diffractor and so in order to obtain larger crystals, trials were set up using variations on Buffer 39 (Table 4.5).

After 3 months crystals had grown with all the conditions but unfortunately they were still too small to mount onto the X-ray diffractor.

0.1 M Na.HEPES pH	PEG400 % (v/v)	Ammonium sulphate conc. (M)
7.2, 7.8	2.0	2.0
7.5	1.6, 1.8, 2.2	2.0
7.5	2.0	1.6, 1.8, 2.2

Table 4.5 The combination of buffers used in crystallisation trials of the mutant pig citrate synthases.

4.3 DISCUSSION

The role of hydrophobic interactions at the subunit interface in pig citrate synthase has been investigated by mutating two polar residues, Ser 139 and Thr 146 into alanine amino acids. These residues are found in helix G at the interface of the protein. This form of mutation has the effect of increasing the alanine content of the helix and also increasing the degree of hydrophobicity at the interface, both features of which have been shown to play a role in the stability of thermophilic proteins [Kotik & Zuber, 1993; Kirino *et al.*, 1994].

The two interface mutants, T146A and S139A/T146A, were created (section 4.2.3), and successfully expressed in the citrate-synthase deficient *E. coli* strain MOB154 (section 4.2.4). The purification of the enzymes was very straightforward, resulting in excellent recovery levels (Table 4.1). There was the possibility that endogenous *E. coli* citrate synthase may have bound to the affinity column, this protein being either the expression product of the 'second' *E. coli* citrate synthase gene

[Patton *et al.*, 1993], or the inactive product of the mutated *E. coli* citrate synthase gene [Ashworth & Kornberg, 1966]. Thus, it was encouraging when a cell extract of *E. coli* MOB154 strain was loaded onto the Matrex™ gel Red A column and no citrate synthase was eluted under identical conditions to those that were employed to purify the pig enzyme (section 4.2.4).

4.3.1 STRUCTURAL CHARACTERISATION

The two subunit interface mutants of pig citrate synthase, T146A and S139A/T146A, were homology modelled prior to the site-directed mutagenesis (section 4.2.2), and analysis of the models proved to be promising, showing that there was no significant conformational change between the wild type citrate synthase structure and the two mutant structures. The CD spectroscopy, on the other hand, revealed slight differences in both the secondary structure content and the tertiary structure of the mutant proteins. Data from the far-UV spectra showed that the helical contents in the two mutants were similar to each other but both slightly less than wild-type (Table 4.4), even though the helical content of the wild-type enzyme had been underestimated somewhat. The underestimation of helical content by CD spectroscopy in pig citrate synthase has been reported previously, and was due to inappropriate choice of a reference value for the mean residue ellipticity [Wu & Yang, 1970]. However, in this case it is thought that the discrepancy may be due to an overestimation of the β–strand content, probably arising as the result of the smooth β–strand-like nature of helices G and M [S. Kelly, pers. comm.]

Data from near-UV spectra (Fig. 4.13) indicated that there were also differences in the tertiary structures of the mutant proteins compared with the wild-type pig enzyme. Again, the differences are only marginal, but this time the variation due to the double mutation was more pronounced than in the single mutant, indicating accumulative changes. These slight conformational disparities between the structures are not really surprising considering that polar residues, found in helices

and involved in intersubunit contacts, have been substituted with the aliphatic and smaller Ala amino acid. It is interesting that the helical content of the mutants should appear to be lower than that in the wild-type pig citrate synthase, and also of equivalent values to each other, since it was hoped that Ala substitutions would increase the stability of Helix G, and that this stability would be proportional to the number of substitutions. However, the differences in helical content may just reflect the gross conformational changes that are also observed with the near-UV spectra (Fig. 4.13), and which could possibly obscure any increased stability of Helix G.

4.3.2 KINETIC CHARACTERISATION

The kinetic characterisation of the wild-type and mutant citrate synthases was carried out (section 4.2.5). The Km values for the two substrates, oxaloacetate and acetyl coenzyme A, are below 10 μM in all three enzymes, and this is in agreement with the Km values previously reported for eucaryal citrate synthases [Weitzman & Danson, 1976]. The Vmax value for mutant S139A/T146A was less than half those for the wild-type enzyme and mutant T146A (140, 370 and 360 μmol/min/mg protein, respectively). There are two possible explanations to account for this difference: Either all of the enzyme is catalytically less efficient, or a portion of the protein is present as inactive molecules, whether denatured or not. The latter theory is unlikely considering that the mechanism of the affinity chromatography employed to purify the proteins requires a coenzyme A binding site, and this can only be formed in the dimeric structure of the active enzyme. The former theory is more feasible, and is probably due to conformational changes induced in the protein as a result of mutating both Ser 139 and Thr 146 residues into Ala amino acids. The results of the structural characterisation of the citrate synthases with CD spectroscopy (4.2.9) may support this explanation, the near-UV spectra showing that the tertiary structure of mutant S139A/T146A is slightly different to that of the wild-type protein and mutant T146A (Fig. 4.13).

4.3.2 THERMAL CHARACTERISATION

From an analysis of the modelled interface mutants (section 4.2.2), a number of intersubunit hydrogen bonds were identified that are absent in the mutant proteins, two in mutant T146A, and four in mutant S139A/T146A. It was realised that these losses could have a detrimental effect upon the thermostability of the mesophilic protein, however, it was hoped that the additional stability resulting from mutating Ala amino acids into helix G and increasing the hydrophobicity at the subunit interface would at least counteract the decrease in bond energy.

The results of the thermal inactivation studies (section 4.2.6) showed that mutant T146A was marginally more stable than mutant S139A/T146A. The energies of activation of the inactivation process (E_a) were very similar, determined as 200 kJ/mol and 189 kJ/mol for the single and double mutants, respectively. More interesting are the results obtained with the wild-type pig citrate synthase thermal inactivation (Fig 4.6(a)), where the inactivation plots were not of the expected linear form. For this reason it was not possible to calculate the value of E_a for the wild-type enzyme for comparison with the E_a values of the mutants. The value of E_a for pig citrate synthase has been cited previously as 331 $\pm$ 22 kJ/mol [Russell, 1994], but this value cannot be used in valid comparisons with those of the mutants because the thermal inactivation of the wild-type enzyme was monitored over a very short time course, only 10 min, and one cannot be sure if the rate of inactivation would not have decreased after an extended time period, as is the case in Fig. 4.6(a). More importantly, the concentration of protein used in the reported thermal inactivations of Russell [1994] is not known. This is a crucial factor, as illustrated by the results of the concentration-dependent inactivation of all three pig citrate synthase species (section 4.2.7). The rate of inactivation was found to increase with a decreasing concentration of enzyme; this is in agreement with the work of McEvily and Harrison [1986]. This group proposes that inactivation of pig citrate synthase is dependant on two consecutive first-order processes. Firstly, the reversible dissociation of dimers, and secondly, the irreversible denaturation of monomers, the rate constants for which are measurable above 1 μM and below

5 nm protein concentrations, respectively. Intermediate concentrations were shown to exhibit a biphasic nature of inactivation.

The results in section 4.2.7, also show a biphasic loss of activity, and more significantly they highlight the differences in the concentration-dependent inactivation of the mutant citrate synthases when compared with the wild-type enzyme. Figures 4.8 and 4.9 illustrate how the rate of inactivation at protein concentrations above 100 nM is dramatically increased in the two mutants compared to the wild-type citrate synthase. Since dimer dissociation is the first step of inactivation, it appears that the mutations at the subunit interface may have affected the dissociation constants (K_d) of the molecules, shifting the equilibria toward the monomeric states. Below 100 nM protein concentrations the rate of inactivation for the three proteins are quite similar, possibly implying that the rate of denaturation of the monomeric species of citrate synthase is unaffected by the mutations in helix G.

From the results of the differential scanning calorimetry (Fig. 4.11) it would appear that the two interface mutants of pig citrate synthase are slightly less thermostable than the wild-type enzyme. Although the difference in melting temperatures is only small, 50°C (wild-type) compared with 47.5°C (mutants), the fact that the peak in heat capacity is shifted by the same amount for both mutants is thought to be significant. The presence of an apparent second transition at 57°C with the wild-type and mutant S139A/T146A citrate synthases may be the result of the independent unfolding of the two domains within each monomer, although, this pattern has not been seen before in pig citrate synthase [D W Hough, pers. comm.]. The reason that the single mutant, T146A, does not possess the second transition is unclear. In addition, with all three citrate synthases there appeared to be no change in the partial C_p of the folded protein compared with the unfolded protein; this could be due to the exothermic process of protein precipitation which occurred in each sample during the DSC analysis.

The results of the thermodynamic analysis of unfolding of citrate synthase due to guanidine denaturation show that there is no clear difference between the unfolding

of the wild-type and the interface mutant enzymes (Figs. 4.14 and 4.15). Unfortunately, the spectroscopic analyses were not detailed enough to enable a determination of ΔG values, and thus need to be repeated more thoroughly before any conclusions can be made concerning the differences in ΔG between the wild-type and the mutant citrate synthases.

Overall, it would seem that increasing the hydrophobicity of Helix G at the subunit interface of pig citrate synthase has resulted in the creation of two proteins that have a slightly reduced stability to temperature and also an increased rate of dimer dissociation. The subtle conformational changes in the secondary and tertiary structures of the proteins may be responsible for the observed differences described. In addition, the kinetics of the double mutant are affected and this could be due to the more pronounced effect of the double mutation upon the tertiary structure of the enzyme. The changes in tertiary structure may be related to the loss of intersubunit hydrogen bonds in both mutants, more so in the double mutant. Much more work needs to carried out with the interface mutants in order to quantify any differences in the kinetic inactivation and thermodynamic unfolding, compared with wild-type pig citrate synthase. Possible future work could include a thorough spectroscopic analysis of the guanidine.HCl denaturation of the mutants in order to determine any changes in the value of ΔG between the wild-type and mutants. The concentration-dependent thermal inactivation of the three protein species also needs to be characterised more fully. Moreover, if the crystal structures of mutants T146A and S139A/T146A are determined they can be used in comparative analyses with the wild-type enzyme structure in order to explain any disparities highlighted between the wild-type and mutant citrate synthases in the above studies.

CHAPTER 5

INVESTIGATING THE EFFECT OF CAVITY MUTATIONS UPON THERMOSTABILITY

5.1 INTRODUCTION

From an analysis involving 121 protein chains it was concluded that cavities are tolerated within a protein structure rather than actively favoured or disfavoured [Hubbard *et al.*, 1994]; this is thought to be because, although the cavities appear to be energetically unfavourable, their presence allows other regions of the molecule to adopt preferred conformations in order to achieve the lowest global free energy [Hubbard *et al.*, 1994]. The general assumption that cavities are destabilising derives from the lower number of van der Waals contacts in them compared to that in a 'fully' packed structure [Pakula & Sauer, 1989]. Site-directed mutagenesis has been used to investigate the effect of both filling cavities, resulting in an increase in the thermostability of the protein [Ishikawa, *et al.*, 1993(b)], and of creating cavities, resulting in a decrease in the thermostability [Eriksson *et al.*, 1992; Ishikawa, *et al.*, 1993(b)]. However, others have found that changes in the packing inside cavities makes little or no difference to the stability of the protein in question [Karpusas *et al.*, 1989]

As described in section 3.3.2, comparative analyses of dimeric citrate synthases from pig, *Tp. acidophilum* and *P. furiosus*, reveal that the volume of the molecule occupied by cavities decreases dramatically as the thermostability of the protein increases. In the present chapter, the role of this feature is investigated by site-directed mutagenesis in order to reveal if the differences in the number of cavities affects the thermostability of the protein.

5.2 RESULTS

5.2.1 MODELLING THE CAVITY MUTANTS OF PIG CITRATE SYNTHASE

In pig citrate synthase residues Ala 118 and Ser 412 line independent cavities. In the *Tp. acidophilum* protein the bulkier Tyr residue occupies the equivalent 3D positions (residue numbers 76 and 356) and thus they eradicate the cavities in the archaeal enzyme structure. Two cavity-filling mutants of pig citrate synthase, a single mutant (S412Y) and a double mutant (A118Y/S412Y), were homology modelled, substituting the residues Ala 118 and Ser 412 with Tyr residues (Fig. 5.1). The modelling was carried out as described in section 2.1 using the open form of pig citrate synthase (1CTS.PDB) as the template.

A superimposition of the two mutants upon the wild type pig citrate synthase showed that there is virtually no change in the α–carbon chains of the mutants. The only slight difference is a shift of 0.6 Å and 0.31 Å at the Cα position of the two substituted Tyr residues when compared with Ala 118 and Ser 412, respectively. Ramachandran plots were identical between the mutants and wild type enzymes indicating that Tyr 118 and Tyr 412 do not adopt strained conformations nor affect the conformation of surrounding residues. In addition, the environment of the mutated residues and also those residues lining the two cavities were found to be unaffected by the amino acid substitutions. The number and nature of hydrogen bonds was maintained between the A118Y mutation and the wild type structure; however, there was a net loss of one hydrogen bond for the S412Y mutation, comprising the loss of two bonds involving Ser 412 but the gain of one hydrogen bond involving Tyr 412 (Table 5.1)

(a)

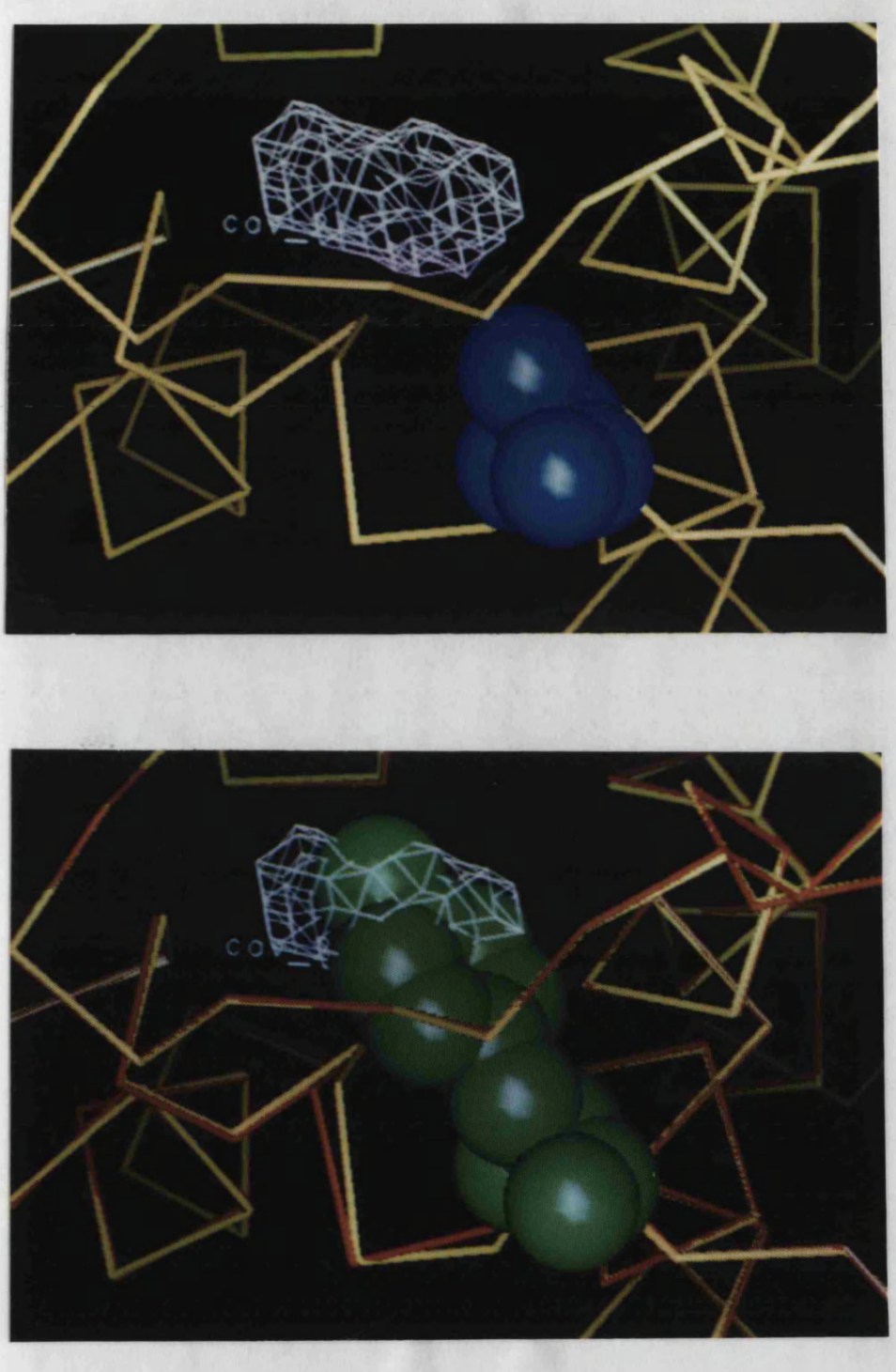
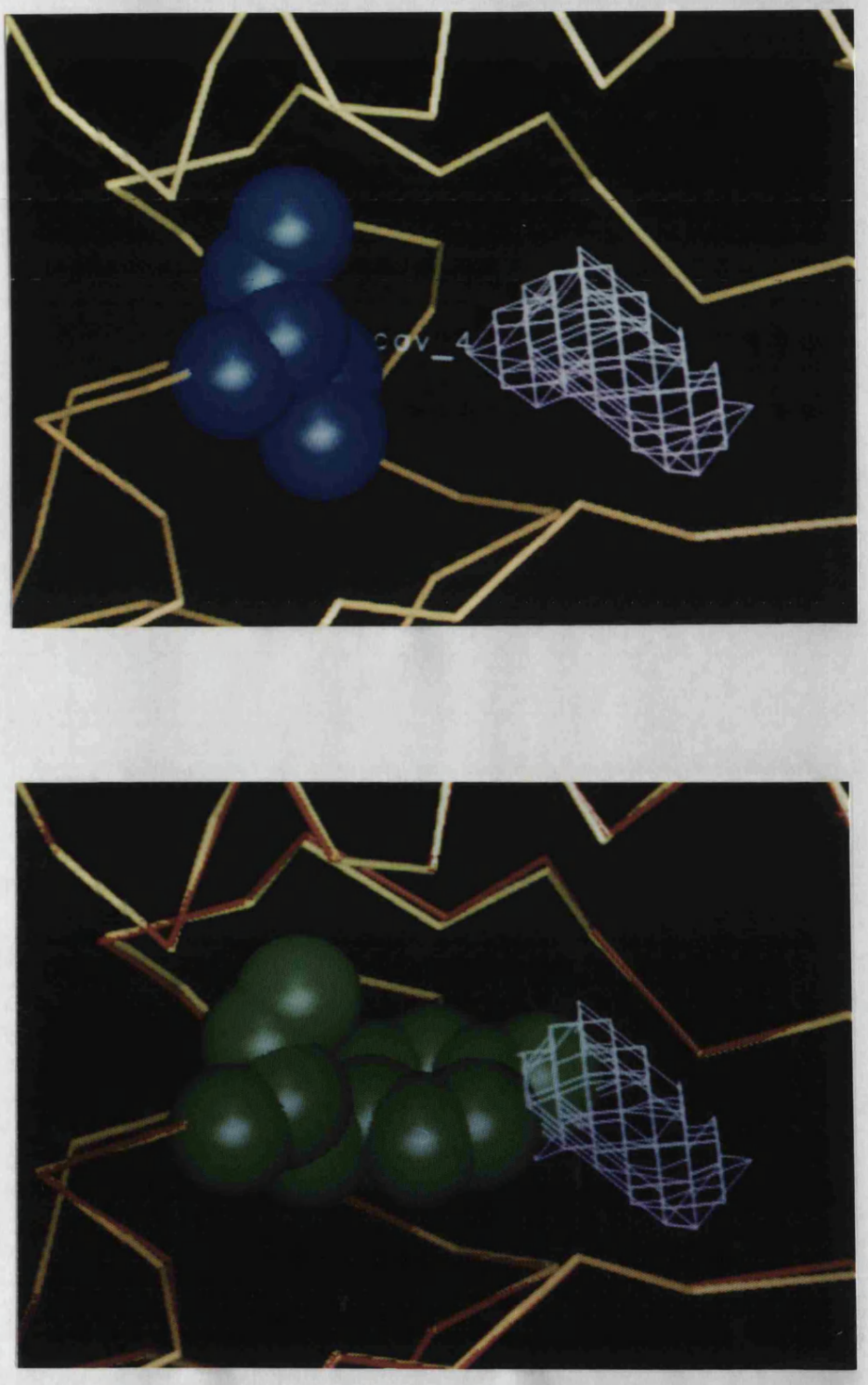


Fig. 5.1 **Cavity-filling mutants of pig citrate synthase.** The α-carbon traces of the wild-type pig citrate synthase (yellow) and the modelled mutants of the protein (red) are shown together with two cavities (hashed), (a) and (b), generated by VOIDOO [Kleywegt & Jones, 1994]. The wild-type residues Ala 118 (a) and Ser 412 (b) are highlighted (blue), as are the equivalent mutant Tyr residues (green).

Fig. 5.1 (b) (See figure legend on previous page)



WILD TYPE PIG CS		CAVITY MUTANTS	
atom #1	atom #2	atom #1	atom #2
Ala 118 N	Trp 114 O	Tyr 118 N	Trp 114 O
Ser 412 OE	Gln 408 O		
Ser 412 OE	Gln 408 NE2		
		Tyr 412 OE	Pro 418 O

Table 5.1 Hydrogen bonds involving Ala 118 and Ser 412 (atom #1) with other residues (atom #2) of wild type pig citrate synthase, and Tyr 118 and Tyr 412 (atom #1) with other residues (atom #2) of the two cavity mutants of pig citrate synthase. Equivalent hydrogen bonds are shown alongside each other.

5.2.2 SITE-DIRECTED MUTAGENESIS OF PIG CITRATE SYNTHASE

The pig citrate synthase gene was mutated, using the pALTER™ method, described in section 2.2.14, to create the single mutant S412Y and the double mutant A118Y/S412Y. The mutagenic oligonucleotides are shown in Fig. 5.2. The double mutant was achieved in one mutagenic reaction using both oligonucleotides simultaneously.

a)

				GCA					
5' (430)	GCA	AAG	AGG	**ATA**	GCT	CTG	CC	(449)	3'
	A115	K116	R117	**Y118**	A119	L120			

b)

				AGC						
5' (1314)	C	ATC	TGG	**ATA**	CGA	GCC	TTA	G	(1333)	3'
		I410	W411	**Y412**	R413	A414	L415			

Fig. 5.2 Mutagenic oligonucleotides to replace (a) Ala 118 with Tyr, and (b) Ser 412 with Tyr. The mutated bases are highlighted in bold, the equivalent wild-type bases are shown above the oligonucleotides and the corresponding residues are shown below.

Mutants were screened by manual sequencing of the double stranded recombinant pALTER™ vector as described in section 2.2.13 (results are shown in Fig. 2 of the Appendix)

5.2.3 EXPRESSION OF CAVITY MUTANTS

Initial expression studies involved the subcloning of S412Y and A118Y/S412Y mutant pig citrate synthase genes into the expression vector pKK223-3 at the *Eco*RI and *Hind*III restriction sites, and their subsequent transformation into the citrate synthase-deficient *E. coli* strain MOB154 (protocols are described in sections 2.2.1 to 2.2.9). Cultures of the transformed *E. coli* were incubated overnight at 37°C and a cell extract was prepared using 1.5 ml of the culture (section 2.3.1). The extracts were assayed for citrate synthase activity at 37°C and 55°C (section 2.3.4), but there was no detectable activity at either temperature. Electrophoresis of the cell extracts and cell debris on 10% SDS-PAGE gels (section 2.3.3) revealed that the two mutant proteins were expressed in an insoluble form that pelleted with the cell debris.

Detailed studies have shown that *in vitro* protein aggregation arises from the association of partially folded intermediates and depends on protein concentration [Jaenicke & Rudolph, 1989; Brems, 1988; Bowden & Georgiou, 1990]. Folding and aggregation are opposing kinetic pathways influenced by the rate of translation, which in turn is governed by the strength of the ribosome binding site, in this case, carried on the expression vectors. The growth temperature of the bacterial cells expressing the foreign protein has also been shown to greatly affect solubility, with the trend towards increased solubility at lower temperatures [Schein & Noteborn, 1988; LaVallie *et al.*, 1993].

In an attempt to produce soluble and active citrate synthase cavity mutants, a variety of expression systems were explored involving expression vectors pKK223-3, pMEX8 and pJLA602, and *E. coli* strains MOB154, DH5α and JM105. The pMEX8 vector is similar to pKK223-3 (described in section 4.3.2), in that both

possess the tac promoter for transcription, and ribosome binding site and termination signals for translation. Insertion of a gene at the *Eco*RI restriction site with its own initiation codon following immediately, results in the highest efficiency of translation. Both vectors code for the βlactamase gene for ampicillin resistance, as does pJLA602 vector. Vector pJLA602 contains the strong major leftward and major rightward promoters from the λbacteriophage for transcription but these are repressed at 28-30°C by the product of the λbacteriophage *c*Its857 gene, which is also present on the vector. Induction is achieved by increasing the incubation temperature to 42°C. The vector also carries the translation initiation region for the *E. coli atpE* gene and six termination signals after the *Bam*H1 restriction site.

For expression in vectors pKK223-3 and pMEX8, the mutant genes, together with wild type pig citrate synthase gene as a control, were inserted at the *Eco*RI and *Hin*dIII restriction sites. For pJLA602, the 5' ends of the genes were inserted at the *Nco*I restriction site, whilst the 3' ends were cloned into the *Bam*HI site. This latter cloning reaction was achieved by digesting the citrate synthase genes with *Hin*dIII, and the vector with *Bam*HI, and then end filling both DNA species so that the 3' end of the genes were blunt-end ligated with the vector. The vectors, with their various inserts, were used to transform the three *E. coli* strains MOB154, DH5α and JM105 (protocols are described in sections 2.2.1 to 2.2.9).

5.2.3.1 EXPRESSION USING PKK223-3 AND PMEX8

Cultures of all the pKK223-3 and pMEX8 carrying strains were incubated overnight at both 30°C and 37°C. Cell extracts were prepared, assayed for citrate synthase activity and electrophoresed, together with cell debris, as described above. For the wild type pig citrate synthase controls, the results of the assays showed that an active citrate synthase was expressed in all the cultures with specific activities ranging from 0.4-26.9 U/mg protein, above any endogenous *E. coli* enzyme (summarised in Table 5.2). The temperature of incubation appeared to have a marked effect upon expression levels only in MOB154 strain, where those cultures incubated at 37°C had a higher specific activity of citrate synthase than the cultures

grown at 30°C. There was a significant difference in the level of expression between strains: using pKK223-3, the amount of enzymic activity decreased in the order MOB154 > DH5α > JM105; with pMEX8, in the order MOB > JM105 > DH5α.

VECTOR	INCUBATION TEMP. (°C)	PCS SPECIFIC ACTIVITY (U/mg) IN *E. COLI* STRAIN		
		MOB154	JM105	DH5α
pKK223-3	30	16	0.76	26.9
	37	24	0.91	25.4
pMEX8	30	19	9.91	0.4
	37	24.5	7.41	0.6

Table 5.2 **Factors affecting the levels of expression of the pig citrate synthase (PCS) gene.** The table shows the levels of wild type pig citrate synthase activity in the cell extracts prepared from different expression system cultures incubated at both 30°C and 37°C. The enzymic activity is expressed in terms of specific activity (U/mg protein).

With the cavity mutants, there was no detectable citrate synthase activity above any endogenous *E. coli* citrate synthase, using the vectors pKK223-3 and pMEX8 and incubations at either 30°C or 37°C. Analysis of the cell extracts and cell debris on SDS-PAGE indicated that there were varied levels of protein expression depending on both strain and incubation temperature (summarised in Table 5.3); however, in all cases, any expressed protein was insoluble (Fig. 5.3).

VECTOR	INCUBATION TEMP. (°C)	EXPRESSION OF PCS IN *E. COLI* STRAIN MOB154	JM105	DH5α
pKK223-3	30	- -	+ +	+ +
	37	+ +	+ +	+ +
pMEX8	30	+ +	+ -	- -
	37	+ +	+ -	- -

Table 5.3 **The expression of cavity mutants of pig citrate synthase from different expression system cultures incubated at both 30°C and 37°C.**

For each strain the first column relates to the single mutant (S412Y) and the second column to the double mutant (A118Y/S412Y). Expression was viewed by SDS-PAGE: expression (+), no expression (-).

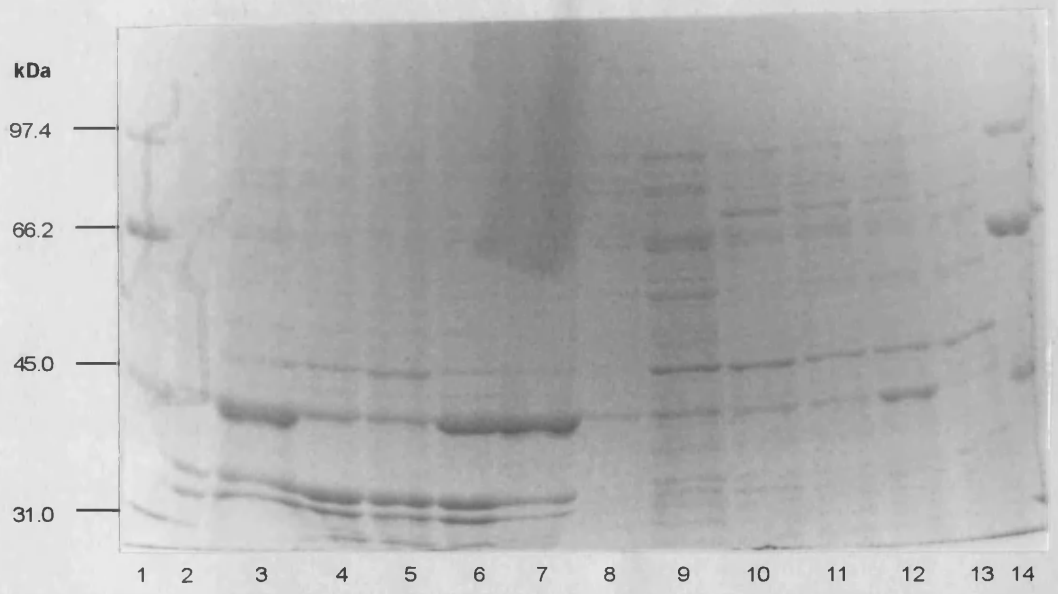


Fig. 5.3 **Expression of the cavity mutants of pig citrate synthase**

(a) SDS-PAGE analysis of the single cavity-filling mutant (S412Y).

Lanes 1 & 14: Protein standards, with molecular weights indicated.

Lanes 2 & 8: Cell debris and cell extracts, respectively, of *E. coli* MOB154 strain carrying the pKK223-3 plasmid-encoded mutant. The cells were grown at 30°C.

Lanes 3 & 9: As for lanes 2 & 8, but grown as 37°C

Lanes 4 & 10: As for lanes 2 & 8, but expressed in *E. coli* JM105 strain

Lanes 5 & 11: As for lanes 4 & 10, but grown as 37°C

Lanes 6 & 12: As for lanes 2 & 8, but expressed in *E. coli* DH5α strain

Lanes 7 & 13: As for lanes 6 & 12, but grown as 37°C

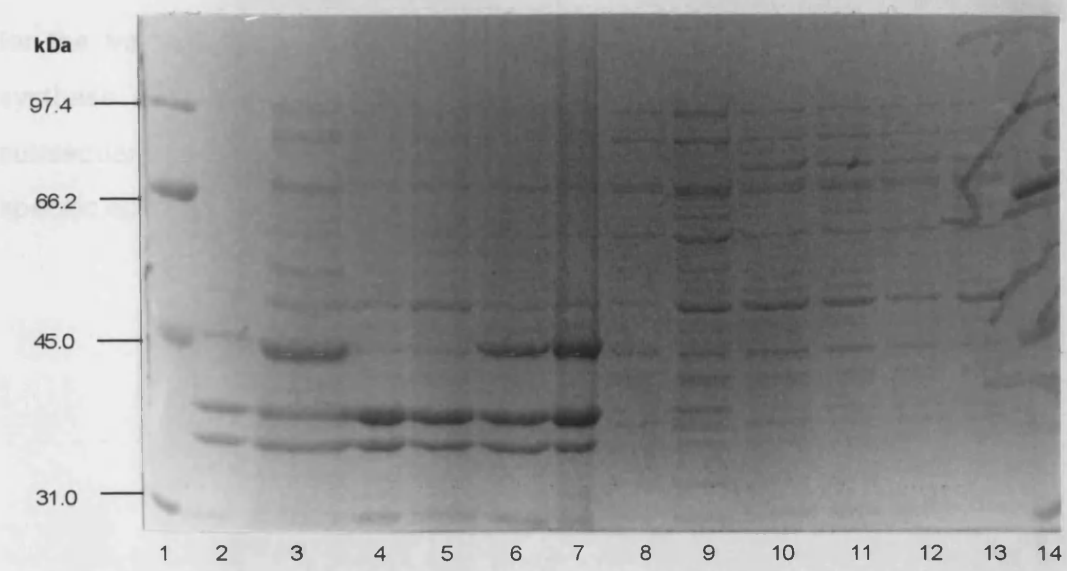


Fig. 5.3 **Expression of the cavity mutants of pig citrate synthase**
(b) SDS-PAGE analysis of the double cavity-filling mutant (A118Y/S412Y)
The lanes are the same as those in Fig. 5.3(a).

5.3.2.2 EXPRESSION USING PJLA602

Expression of wild type pig citrate synthase in the vector pJLA602 was investigated before assessing the expression of the cavity mutants using this vector. Cultures of the three *E. coli* strains transformed with the recombinant pJLA602 vector, and untransformed as control, were grown overnight at 30°C, followed by both (i) induction at 42°C for 75 min then incubation at 37°C for 3 h, and (ii) induction at 42°C over a time course of 30 min, 1 h, 2 h and then incubation at 37°C overnight. Prior to the incubation overnight at 37°C, a 1.5 ml volume of the culture was centrifuged at 13,000 x *g* for 15 s and the cell pellet resuspended in 8 ml LB medium supplemented with ampicillin (section 2.2.1).

Cell extracts of the cultures were prepared, assayed for citrate synthase activity and electrophoresed, together with cell debris, as described above. For the control, the results of the assays show that there was no change in the endogenous levels of *E. coli* citrate synthase in the untransformed bacterial cultures. On the other hand,

for the transformed cells it appeared that the level of expression of pig citrate synthase was dependent upon both the lengths of induction at 42°C and the subsequent incubation at 37°C, such that as these time periods were extended the specific activity of pig citrate synthase increased (Table 5.4).

LENGTH OF INDUCTION (h)	PCS SPECIFIC ACTIVITY (U/mg) IN *E. COLI* STRAIN		
	MOB154	JM105	DH5α
0.5	0.10	0.06	0.14
1	0.15	0.21	0.19
2	0.85	1.51	0.54

Table 5.4 **Expression of wild-type citrate synthase with pJLA.** Table showing the level of wild type pig citrate synthase (PCS) activity in cell extracts of three *E. coli* strains, MOB154, JM105 and DH5α carrying the pJLA602 plasmid-encoded pig enzyme gene, as a function of the length of induction at 42°C. All were followed by a 19 h incubation at 37°C. The enzymic activity is expressed as specific activity in U/mg protein. There was no enzyme activity, above endogenous *E. coli* citrate synthase activity, in those cells induced at 42°C for 75 min, followed by an incubation at 37°C for 3 h.

Expression levels of wild type pig citrate synthase were low, when compared with those using the vectors pKK223-3 and pMEX8 (Table 5.2), as judged by SDS-PAGE of the cell extracts (Fig 5.4). It would also seem that a portion of the wild type protein was insoluble since there appeared to be an additional band at approximately 49 kDa in the cell debris on the SDS-PAGE gels (Fig. 5.4).

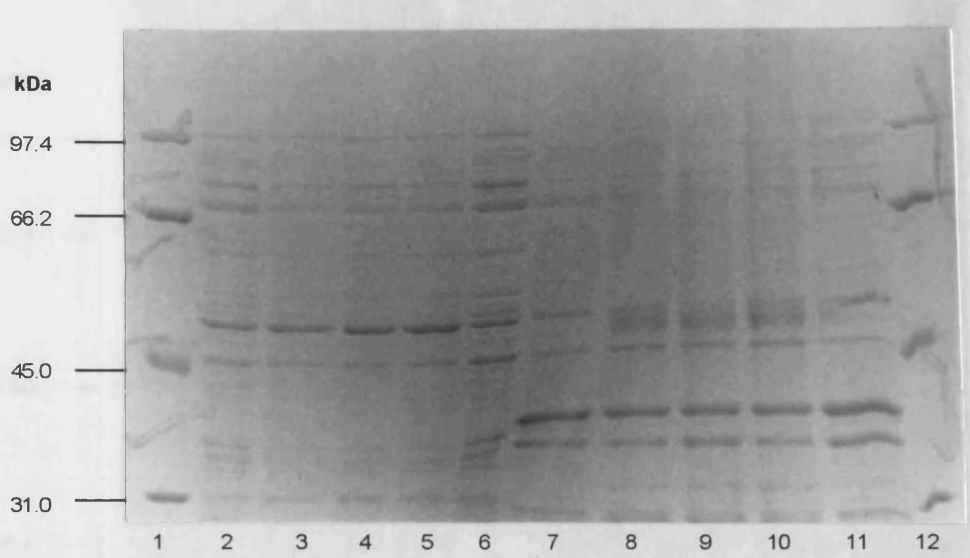


Fig. 5.4 **SDS-PAGE analysis of the expression of wild-type citrate synthase with pJLA602.**
Lanes 1 & 12: Protein standards, with molecular weights indicated
Lanes 2 & 7: Cell extract (CE) and cell debris (CD), respectively, of *E. coli* strain JM105
Lanes 3 & 8: CE's and CD's, respectively, of *E. coli* strain JM105 carrying the pJLA602 plasmid-encoded wild type citrate synthase gene. The cells were induced at 42°C for 30 min followed by incubation at 37°C for 19h
Lanes 4 & 9: As for lanes 3 & 8, but induced for 1 h at 42°C
Lanes 5 & 10: As for lanes 3 & 8, but induced for 2 h at 42 °C
Lanes 6 & 11: As for lanes 3 & 8, but induced for 75 min at 42°C followed by incubation for 3 h at 37°C.

Conditions employed for the expression of the two cavity mutants of pig citrate synthase using the pJLA602 vector involved both (i) induction at 42°C for 90 min followed by incubation at 37°C for 3 h, and (ii) induction at 42°C for 3 h then incubation at 37°C overnight. As with the expression of wild-type protein, bacteria were transferred into fresh media supplemented with antibiotic prior to the overnight incubation. Cell extracts of the cultures were prepared, assayed for citrate synthase activity and electrophoresed, together with cell debris, as described above. There appeared to be no active plasmid-encoded citrate synthase expressed for either mutant in all three *E. coli* strains under the two different induction conditions. In addition, when assessed on SDS-PAGE (Fig. 5.5), the levels of mutant protein expression seemed to be very low, as with the wild type expression, when compared to those levels expressed using pKK233-3 and pMEX8 vectors (Fig. 5.3).

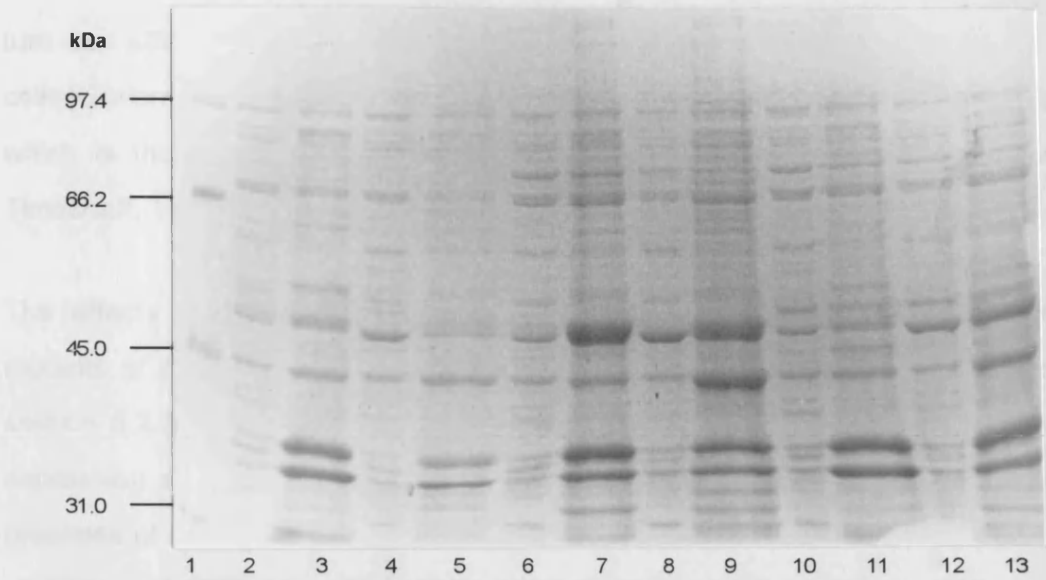


Fig. 5.5 **Expression of the citrate synthase cavity mutants with pJLA602.**
The results of the SDS-PAGE analysis were the same for both cavity mutants and so only the results for the S412Y mutant have been shown.
Lane 1: Protein standards, with molecular weights indicated
Lanes 2 & 3: Cell extract (CE) and cell debris (CD), respectively, of *E. coli* JM105 strain carrying the pJLA602 plasmid-encoded cavity mutant of citrate synthase, S412Y. The cells were induced at 42°C for 1.5 h, then incubated at 37°C for 19 h.
Lanes 4 & 5: As for lanes 2 & 3, but induced at 42°C for 3 h
Lanes 6 & 7: As for lanes 2 & 3, but expressed in *E. coli* MOB154 strain
Lanes 8 & 9: As for lanes 6 & 7, but induced at 42°C for 3 h
Lanes 10 & 11: As for lanes 2 & 3, but expressed in *E. coli* DH5α strain
Lanes 12 & 13: As for lanes 10 & 11, but induced at 42°C for 3 h

5.2.4 INVESTIGATING THE EFFECT OF ORGANIC COMPOUNDS ON PROTEIN EXPRESSION.

It has been previously reported that non-metabolisable sugars and polyalcohols, such as raffinose, sucrose, sorbose and sorbitol, and 'compatible osmolytes', such as betaine, an amino acid derivative, can have a profound effect upon the levels of expression of soluble protein [Bowden & Georgiou, 1990; Blackwell & Horgan, 1991]. The sugars increase the osmotic pressure of the growth medium, which in

turn can affect the growth rate and expression of certain genes within the *E. coli* cells [Gutierrez *et al.*, 1987]. The osmotic stress facilitates the uptake of the betaine which is thought to preferentially hydrate and stabilise the protein [Arakawa & Timasheff, 1985].

The effects of sorbitol, sucrose and betaine upon production of soluble cavity mutants of pig citrate synthase were investigated. From the results of studies in section 5.2.3, the vector pKK223-3 and *E. coli* strain DH5α were chosen as the expression systems with which to work. Bacterial cultures were incubated in the presence of a range of betaine, sorbitol and sucrose concentrations at temperatures of 30°C and 37°C (see Table 5.5). Cell extracts of 1.5 ml culture volume were prepared in the presence and absence of 1 M betaine (section 2.3.1). The extracts were assayed for citrate synthase activity at both 37°C and 55°C (section 2.3.4) and electrophoresed, together with cell debris, on 10% SDS-PAGE gels (2.3.3).

BETAINE CONC. (mM)	SORBITOL CONC. (M)	SUCROSE CONC. (M)
2.5	0.5 & 1.0	
5.0	0.5 & 1.0	
2.5		0.5 & 1.0

Table 5.5 Combination of concentrations of betaine together with either sorbitol or sucrose, used in the expression studies. Those cultures containing 0.5 M sorbitol/sucrose were incubated for 20 h at 30°C and 17 h at 37°C. Those cultures containing 1.0 M sorbitol/sucrose were incubated for 30 h at 30°C and 24 h at 37°C. It was not possible to prepare cell extracts from the cultures containing 1 M sucrose due to difficulties with the sonication of the cells.

The results of the assays showed that there was no active citrate synthase expressed above the endogenous levels of citrate synthase in the *E. coli* DH5α strain, the specific activity of which remained equal to cells grown in the absence of

the alcohol, sugar and betaine. Sonication in the presence of 1 M betaine had no effect upon the levels of enzyme activity in the cells. Electrophoresis of the extracts and cell debris indicated that the sorbitol/betaine and sucrose/betaine did have a positive effect upon the level of production of soluble protein, as seen by the addition of a band at approximately 49 kDa in the cell free extract on the gels (Fig. 5.5). Indeed, for those cultures grown with 1.0 M sorbitol at 30°C, most of the protein appeared to be expressed in a soluble form.

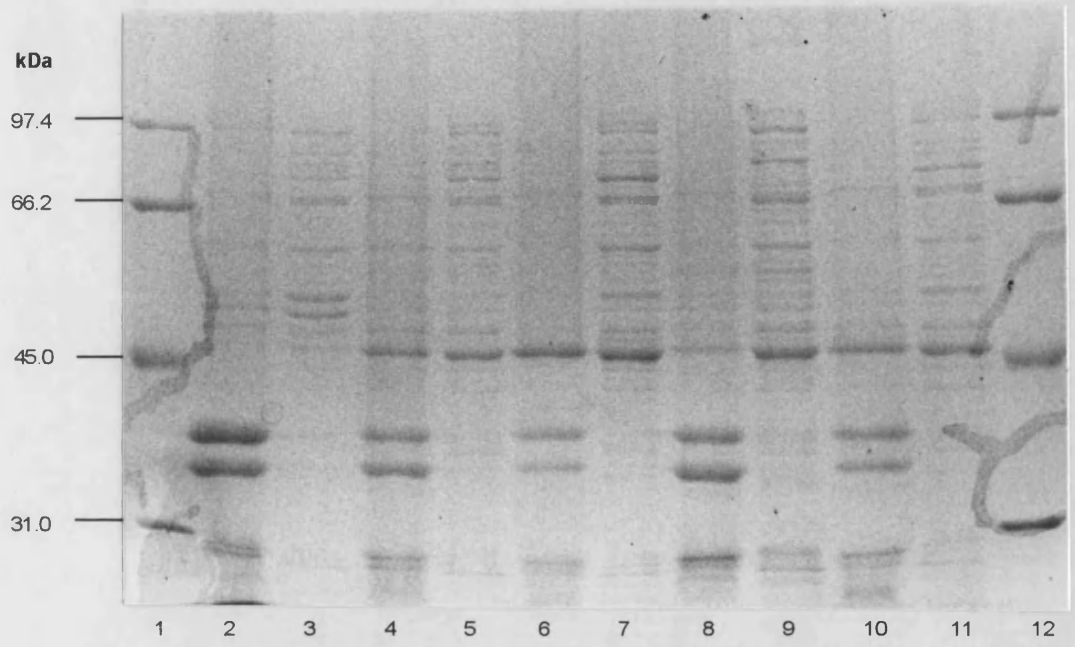


Fig. 5.5(a) **SDS-PAGE analysis illustrating the effects of sorbitol and betaine upon expression of soluble protein.**

The results of the SDS-PAGE analysis were the same for both cavity mutants and at both 2.5 mM and 5.0 mM betaine concentrations, thus only those for mutant S412Y at 2.5 mM betaine have been shown.

Lanes 1 & 12: Protein standards, with the molecular weights indicated

Lanes 2 & 3: Cell debris (CD) & cell extract (CE), respectively, of *E. coli* strain DH5α, grown at 37°C in the presence of 2.5 mM betaine and 0.5 M sorbitol

Lanes 4 & 5: CD & CE of *E. coli* strain DH5α carrying the pKK223-3 plasmid-encoded cavity mutant of citrate synthase, grown in the presence of 2.5 mM betaine and 0.5 M sorbitol, at 30°C

Lanes 6 & 7: As for lanes 4 & 5, but grown at 37°C

Lanes 8 & 9: As for lanes 4 & 5, but grown in the presence of 2.5 mM betaine and 1.0 M sorbitol

Lanes 10 & 11: As for lanes 8 & 9, but grown at 37°C

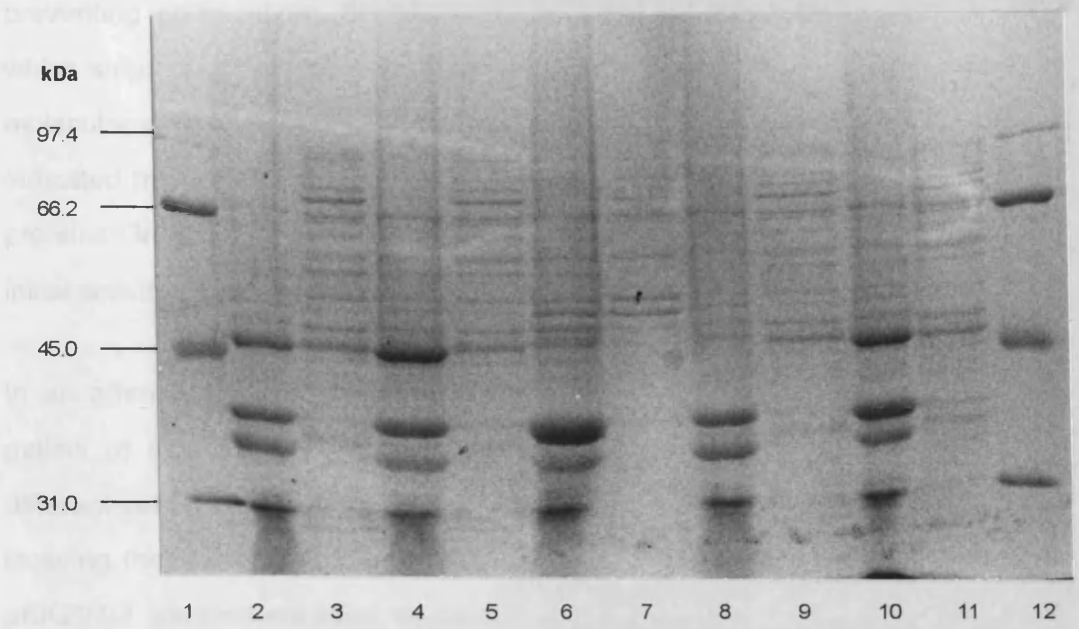


Fig. 5.5 (b) **SDS-PAGE analysis illustrating the effects of sucrose and betaine upon expression of soluble protein.**

Lanes 1 & 12: Protein standards, with the molecular weights indicated

Lanes 2 & 3: Cell debris (CD) & cell extract (CE), respectively, of *E. coli* strain DH5α carrying the pKK223-3 plasmid-encoded cavity mutant (S412Y) of citrate synthase, grown in the presence of 2.5 mM betaine and 0.5 M sucrose, at 30°C

Lanes 4 & 5: As for lanes 2 & 3, but grown at 37°C

Lanes 6 & 7: CD & CE, respectively, of *E. coli* strain DH5α, grown at 37°C in the presence of 2.5 mM betaine and 0.5 M sucrose

Lanes 8 & 9: As for lanes 2 & 3, but carrying the A118Y/S412Y mutant instead

Lanes 10 & 11: As for lanes 8 & 9, but grown at 37°C

5.2.5 DENATURATION AND REFOLDING OF CITRATE SYNTHASE.

The reactivation of guanidine-denatured citrate synthase has been reported previously with a recovery rate of 25% initial activity when the guanidine.HCl was removed by dilution [D. W. Hough, pers. comm.] and 66% when diluted in the presence of polyoxyethylene 10 lauryl (POE(10)L) and β-cyclo-dextrin (β-CD) [Rozema & Gellman, 1995]. The latter method involves the formation of a complex between the POE(10)L, a non-ionic detergent, and the denatured protein thus

preventing aggregation. Protein folding is then initiated by the addition of β-CD which strips away the detergent to allow correct refolding. In this manner the low molecular weight agents are acting as artificial chaperones. Other refolding studies indicated that dilution of guanidine.HCl in the presence of the bacterial chaperone proteins GroEL and GroES together with ATP, resulted in only a 40 % recovery of initial activity [Buchner et al., 1991].

In an attempt to gain soluble and active cavity mutants of pig citrate synthase, pellets of the aggregated protein were treated with guanidine.HCl prior to two different refolding experiments being carried out. The initial experiment involved isolating the cell debris from 1.5 ml cultures of *E. coli* MOB154 strain carrying the pKK223-3 plasmid-encoded mutants, S412Y and A118Y/S412Y (section 2.3.1). The pellets of debris and aggregated protein were resuspended in 6 M guanidine.HCl (Ultrapure, Sigma) and 20 mM dithiothreitol (DTT) and incubated for 15 min at room temperature, after which the guanidine.HCl was diluted 67 fold into 2TE8. This refolding solution was allowed to sit for 2 h during which time aliquots were assayed for citrate synthase activity at 37°C (section 2.3.4). Wild type citrate synthase (1 μg/ml in the refolding solution) was used as a control both under denaturing and non-denaturing conditions in order to assess the percentage recovery. After 30 min refolding there was no further increase in the level of recovery of renatured wild type citrate synthase, which remained at 25% initial activity. After 2 h the activity of the non-denatured citrate synthase had dropped to 95% of its initial value; however, after the same time period the mutant citrate synthases had still not developed any enzymic activity.

The second experiment involved purifying the inclusion bodies containing the aggregated protein, using a modified method of Fischer *et al.* [1992]. A cell extract was prepared from the bacteria of a 200 ml culture volume of *E. coli* strain DH5α carrying the pKK223-3 plasmid-encoded mutant citrate synthase genes (section 2.3.1). The cell debris pellet was washed with 100 ml of 50 mM phosphate buffer, pH 7.4, 0.15 M NaCl, 1 mM EDTA, followed by agitation in 100 ml of 2 M urea, in the above buffer, for 20 min. The purified inclusion bodies were then pelleted at 4,000 x *g* for 30 min at 4°C and then resuspended in 50 μl MilliQ water. A series of

samples were set up that contained the equivalent of 0.01, 0.1, 1.0, 10 and 20 mg of pure inclusion bodies, 6 M guanidine.HCl and 36 mM DTT. These samples were left to denature for 1 h at room temperature, after which they were diluted 50 fold into a buffer that contained 146 mM Tris.HCl, pH 8.0, 0.742 mM EDTA, 0.582 mM POE(10)L. Following another one-hour incubation, the samples were further diluted in a ratio of 7:3 with a solution of 5.3 mM β-CD. The final concentration of reagents in this refolding mix was 84 mM guanidine.HCl, 0.49 mM DTT, 100 mM Tris.HCl, 0.5 mM EDTA, 0.4 mM POE(10)L and 1.6 mM β-CD. Aliquots of the mix were assayed for citrate synthase activity (section 2.3.4) over a time course extending to 15 h, time zero being taken at the point of β-CD addition. Controls included, (i) inclusion bodies (1 mg equivalent) denatured but refolded in the absence of POE(10)L and β-CD, (ii) wild type citrate synthase (0.34 μg/ml final concentration in the refolding mix) untreated with guanidine.HCl but incubated with the POE(10)L and β-CD, (iii) wild type enzyme denatured and refolded in the presence and (iv) absence of POE(10)L and β-CD. For the wild type citrate synthase controls, the results showed that the native enzyme decreased to 64% of initial activity over the 15 h time course, 66% activity was regained after refolding the denatured wild type protein in the presence of POE(10)L and β-CD after 15 h, whilst only 28% was regained in the absence of POE(10)L and β-CD, all of which was present at time zero. Unfortunately there was no active citrate synthase detected at any time in any of the refolding mixtures involving the cavity mutants. An attempt was made to precipitate any soluble protein in the refolding mix, following pelleting of all insoluble matter at 13,000 x *g*, using both 10% (w/v) trichloro acetic acid and 80% saturated solution of ammonium sulphate, but with no success.

5.3 DISCUSSION

The work carried out in this chapter has highlighted the difference between theory and practice of site-directed mutagenesis. Modelling of the cavity-filling mutants of pig citrate synthase gave promising results that suggested that substituting Ala 118 and Ser 412 with Tyr residues would not greatly affect the thermodynamic stability of the proteins, the only significant alteration being the net loss of one hydrogen bond where the Ser 412 to Tyr mutation was concerned (5.2.1). However in reality, such seemingly harmless mutations have resulted in the expression of a protein that primarily aggregates in solution (5.2.3). It is likely that the mutations have hindered the minor structural adjustments of intermediates that are required to take place along the folding pathway resulting in aggregation of partially-folded peptides. The Ala 118 and Ser 412 occur at the C-terminal end of helices E and S, respectively, in pig citrate synthase, whilst in the *Tp. acidophilum* enzyme Tyr 56, equivalent to Ala 118, is found in the loop between helices E and F. Tyrosine residues are known to destabilise α-helical structures, more so than Ser, due to the restrictive conformational freedom of the Cβ substituted side chain and also their ability to hydrogen bond to the main chain in the unfolded state [Horovitz et al., 1992; Ptitsyn & Finkelstein, 1983]; therefore, their substitution into the secondary structure of pig citrate synthase maybe responsible for the incorrect folding of the protein.

Attempts to produce soluble protein were initially unsuccessful (5.2.3). Expression vector systems employing different translation initiation and termination signals were tried in three different host strains of *E. coli*, but to no avail. It may be that the rate of protein expression using vectors pKK223-3 and pMEX8 was too high, and so the kinetic process of aggregation prevailed above the folding process. This is in spite of growing the cells below their optimum growth temperature and purposely not inducing expression. On the other hand, expression levels of citrate synthase using vector pJLA602 were particularly low and this may reflect the need to optimise the induction and incubation conditions with this vector. It was interesting to note that there was differential expression of the citrate synthase for each vector in the three different *E. coli* strains, and that this did not reflect the supplier's recommended host

strain. From this point of view, the results of the expression system studies were useful when choosing a system to use in subsequent experiments.

Soluble protein of the cavity mutants was eventually produced with the use of sorbitol, a polyhydroxy alcohol, sucrose and betaine, an amino acid derivative (5.2.4). The betaine may have stabilised the partially-folded intermediates and so prevented aggregation. However, the two mutants were inactive and therefore one must assume that either the soluble proteins were incorrectly folded, thus producing inactive enzymes, or that they were folded correctly but unable to carry out the citrate synthase reaction. The latter could be due to restrictions upon the conformational movements required for activity of the dimeric citrate synthase, where the small domain rotates 19° with relation to the large domain upon substrate binding [Remington *et al.*, 1982].

Denaturing and refolding experiments were conducted in order to encourage the aggregated citrate synthase mutants to fold correctly (5.2.5). These experiments, making use of low molecular weight agents, POE(10)L and β-CD as artificial chaperones, were very successful for wild type citrate synthase, recovering 66% of initial activity, or 100% of residual activity. However, the conditions were unsuccessful for refolding of active mutants and it appeared that no protein was made soluble during the refolding step, since none precipitated with TCA nor ammonium sulphate.

Further analysis of the cavity-filling mutants needs to be carried out to obtain active protein. There may be other expression systems that can be employed together with the use of sorbitol and betaine in the growth medium that will be more effective than those tried in sections 5.2.3 and 5.2.4. Expression of a foreign gene as a fusion protein tends to be successful for the production of both soluble and active proteins [LaVallie *et al.*, 1993]. Unfortunately, this type of system is unsuitable for use with mutant citrate synthases since the linker residues remaining after cleavage of the fusion could affect the stability of the citrate synthase. On the other hand, the denaturing and refolding of the insoluble or soluble forms of the two mutants could be optimised, maybe using urea or NaCl, instead of guanidine, in the denaturing

step. Refolding studies were not carried out on the soluble mutant proteins expressed in the presence of sorbitol/sucrose and betaine (section 5.2.4) due to the high concentration of endogenous *E. coli* proteins that could interfere with the refolding process. Nevertheless, these experiments could be conducted denaturing either the cell extracts or, more preferably, a purified sample of the cavity mutants.

If soluble and active protein of the pig citrate synthase cavity mutants is obtained then characterisations of the kinetics, thermostability and structure of the enzymes can be carried out, as in Chapter 3 with the subunit interface mutants, in order to investigate the effect of reduction in cavity size.

CHAPTER 6

THE ROLE OF THE N-TERMINUS IN CITRATE SYNTHASE

6.1 INTRODUCTION

There are two forms of the citrate synthase monomer, one being approximately 40 residues longer at the N-terminus than the other. At the time when the experiments described in this chapter were carried out it was thought that these N-terminal residues may play a role in the thermostability of citrate synthase, as described in section 3.3.2.1, since all the archaeal citrate synthases were of the shorter form, as was that from a thermotolerant *Bacillus* species. However, more-recently elucidated citrate synthase sequences from mesophilic organisms have also been of the shorter form and so it is now believed that the N-terminal arm could represent a phylogenetic marker instead. Nevertheless, the absence of the terminal residues makes citrate synthase from *Tp. acidophilum* a more compact and less flexible structure, which could play some role in the thermostability of the protein, and so this has been tested by site-directed mutagenesis.

6.2 RESULTS

6.2.1 HOMOLOGY MODELLING THE TRUNCATED PIG CITRATE SYNTHASE

A mutant of pig citrate synthase, with the first 36 N-terminal residues removed, was homology modelled using the 3D structure of the open form of pig citrate synthase (1CTS.PDB) as the template (section 2.1).

The DSSP program was used to identify residues in the monomer that were exposed to solvent as a result of removing the N-terminus of the structure, as judged by a change in the surface accessibility of at least 5 $Å^2$. There are 24 residues that became exposed, half of which are hydrophobic, six charged and the remaining six polar. The environment of the exposed residues was checked and it appeared that none were in an unsuitable environment. In comparison, in the wild-type citrate synthase monomer, residues from Lys 22 to Val 32 were shown to have the lowest suitability to their environments in the molecule. In the wild-type enzyme there are 31 hydrogen bonds involving residues of the N-terminal arm, 27 of these bonds are within the terminus structure mostly along the α-helix; the remaining four bonds are to other non-contiguous residues of the subunit. All these hydrogen bonds are absent in the truncated mutant.

Upon dimer formation, the majority of residues 22-36 are buried through packing with the equivalent residues in the other subunit. This packing is complemented by a significant increase in the suitability of the environment of these residues. There are six hydrogen bonds, formed between the N-terminal residues of the two subunits, and a further ten hydrogen bonds formed with other residues of the second subunit, all of which are absent in the dimer of the truncated mutant citrate synthase.

In summary, analysis of the model of the truncated pig citrate synthase showed that the structure was thermodynamically stable but with a significant reduction in the number of hydrogen bonds in both the monomer and dimer of the protein. To compensate for this loss of bonds, it is thought that the removal of the N-terminal arm from the protein will result in a less flexible structure, particularly since the arm is one of the most highly flexible regions of the protein [Remington *et al.*, 1982].

6.2.2 DELETING THE N-TERMINUS FROM PIG CITRATE SYNTHASE

The first 36 N-terminal residues of pig citrate synthase were removed using the polymerase chain reaction (PCR). Plasmid DNA (clone pCS4), 10 μg was incubated with 1 μM primers (Fig. 6.1), 300 μM of each deoxynucleoside triphosphate and $Vent_R$™ buffer (10 mM KCl, 20 mM Tris.HCl pH8.8, 10 mM $(NH_2)_2SO_4$, 2 mM $MgSO_4$ 0.1 %(v/v) Triton X-100) at 95°C for 5 min, after which 1 U of $Vent_R$™ was added. The PCR was continued for 30 cycles, melting at 94°C for 90 s, annealing at 37°C for 90 s and extending at 72°C for 90 s. A final incubation was carried out at 72°C for 10 min. Reaction controls involved omitting each primer in turn and also omitting the template plasmid DNA. The PCR was analysed on a 1 % (w/v) agarose gel (section 2.2.5) (Fig. 6.2) and then the amplified DNA was purified from a 1 % (w/v) LMP agarose gel (section 2.2.6). The DNA was ligated into M13mp19 at the *Eco*RI and *Hind*III restriction sites, and the resultant plasmid was used to transfect *E. coli* strain DH5αF' (protocols are described in sections 2.2.4 to 2.2.10). Single-stranded DNA was isolated (section 2.2.11) and then sequenced manually to check that translation would start with the introduced Met followed by residue Thr 37 (section 2.2.13) (results are shown in Fig. 3 of the Appendix).

Forward PCR primer:

5'	(176)	AC	ACC	GTG	GTG	***AAT***	***TC***A	ATG	ACT	GTG	GAC	(204) 3'
						*Eco*RI		start	T37	V38	D39	

Fig. 6.1 **Primer used in PCR to remove first 36 N-terminal residues from the pig citrate synthase gene.** The forward primer introduces an *Eco*RI site (*GAATTC*) and a translation start codon (ATG). All the mutations are shown in bold, the base numbers are in brackets and the amino acids are shown below their codon. The reverse primer is the same as that used in section 4.2.1 (Fig. 4.1).

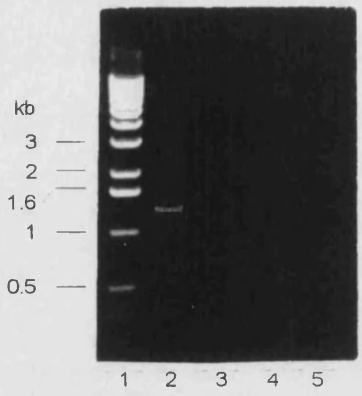


Fig. 6.2 **PCR to create a truncated pig citrate synthase.**

Lane 1: 1 kbase DNA size markers

Lane 2: PCR product

Lanes 3-5: Control reactions omitting the forward primer (3) the reverse primer (4) and the template DNA (5)

6.2.3 EXPRESSION OF THE TRUNCATED CITRATE SYNTHASE MUTANT

The truncated mutant of pig citrate synthase was ligated into the expression vector pKK223-3 (described in section 4.2.3), at the *Eco*RI and *Hind*III restriction sites, and this was used to transform the citrate synthase-deficient *E. coli* strain MOB154 (protocols are described in sections 2.2.1 to 2.2.9). A cell extract was prepared from an overnight culture of the transformed *E. coli* (section 2.3.1), and this was assayed for citrate synthase activity at 37°C and 55°C (section 2.3.4). The extract, together with the cell debris, were electrophoresed on a 10% SDS-PAGE gel (section 2.3.3).

This initial expression study showed that there was no detectable citrate synthase activity in the cell extract nor any expressed protein detectable on SDS-PAGE. Thus, a range of different expression systems were tested with the truncated mutant, equivalent to those that were used in section 5.2.3.1 with the cavity mutants of pig citrate synthase. The mutant gene was ligated into a second expression vector, pMEX8 (described in section 5.2.3), at the *Eco*RI and *Hind*III restriction sites, and both recombinant vectors were then used to transform the three *E. coli* strains

MOB154, DH5α and JM105 (protocols same as above). The bacteria were incubated at both 30°C and 37°C for 19 h, after which cell extracts were prepared, assayed for citrate synthase activity and electrophoresed, together with cell debris, on 10% SDS-PAGE gels, as described above. Wild-type citrate synthase was expressed under identical conditions and the results of this control, described in section 5.2.3.1, show that an active, soluble citrate synthase was produced and that it was detectable in the cell extract on SDS-PAGE gels as an additional band at 49 kDa. Unfortunately, with the truncated mutant, there was no active citrate synthase, above any endogenous *E. coli* enzyme activity, detected in any of the cell extracts grown under all the conditions. SDS-PAGE gels showed that there was expression of the mutant protein but only using the vector pKK223-3 in the *E. coli* strain DH5α, and that the protein was insoluble (Fig. 6.3, lanes 6 & 7).

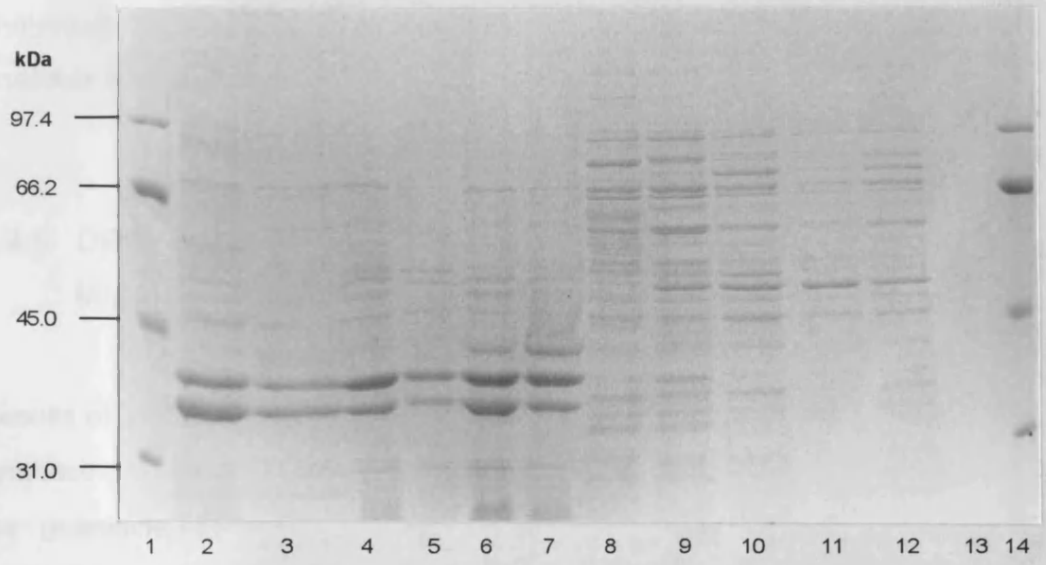


Fig. 6.3 **SDS-PAGE analysis of the expression of the truncated mutant of pig citrate synthase.**
Lanes 1 & 14: Protein standards, with molecular weights indicated.
Lanes 2 & 8: Cell debris and cell extracts, respectively, of *E. coli* MOB154 strain carrying the pKK223-3 plasmid-encoded mutant enzyme. The cells were grown at 30°C.
Lanes 3 & 9: As for lanes 2 & 8, but grown as 37°C
Lanes 4 & 10: As for lanes 2 & 8, but expressed in *E. coli* JM105 strain
Lanes 5 & 11: As for lanes 4 & 10, but grown as 37°C
Lanes 6 & 12: As for lanes 2 & 8, but expressed in *E. coli* DH5α strain
Lane 7: As for lane 6, but grown as 37°C

6.2.4 INVESTIGATION OF THE EFFECT OF ORGANIC COMPOUNDS ON PROTEIN EXPRESSION

In an attempt to express an active, soluble truncated mutant of pig citrate synthase, the effects of differing concentrations of sorbitol, sucrose and betaine upon the expression levels of the protein was investigated, as described in section 5.2.4. The mutant gene was produced using the vector pKK223-3 in the *E. coli* strain DH5α, and the conditions employed are shown in Table 5.5. Cell extracts were prepared from cultures that had been incubated at both 30°C and 37°C. The extracts were made in the presence and absence of 1 M betaine, as in section 5.2.4, and then assayed for citrate synthase activity at 37°C and 55°C (section 2.3.4). The extracts and cell debris were electrophoresed on 10% SDS-PAGE gels (section 2.3.3). The results of the assays show there was no detectable citrate synthase activity, above any endogenous *E. coli* enzyme activity, in any of the extracts. Moreover, from analysis by SDS-PAGE, it appears there is mutant protein expression, in both insoluble and soluble forms, only with the 0.5 M sorbitol concentration (Fig. 6.4).

6.2.5 DENATURATION AND REFOLDING OF THE TRUNCATED MUTANT.

Results of guanidine denaturation and refolding experiments on the wild-type citrate synthase (section 5.2.5) revealed that 66 % of initial activity can be recovered when the guanidine.HCl is diluted out in the presence of the non-ionic detergent POE(10)L, and β-cyclodextrin. Therefore, denaturation and refolding experiments were carried out on the aggregated truncated mutant of pig citrate synthase in an attempt to refold the expressed protein into a soluble and active form. The conditions used were the same as those employed for the two experiments on the cavity mutants of citrate synthase, described in section 5.2.5. Both cell debris pellets and purified inclusion bodies were treated with guanidine.HCl, which was then removed by dilution in the presence and absence of POE(10)L and β-CD. The truncated mutant did not refold into a soluble form when the cell debris pellet was

treated with guanidine.HCl, nor when the purified inclusion bodies were treated and the refolding carried out in the presence and absence of POE(10)L and β-CD.

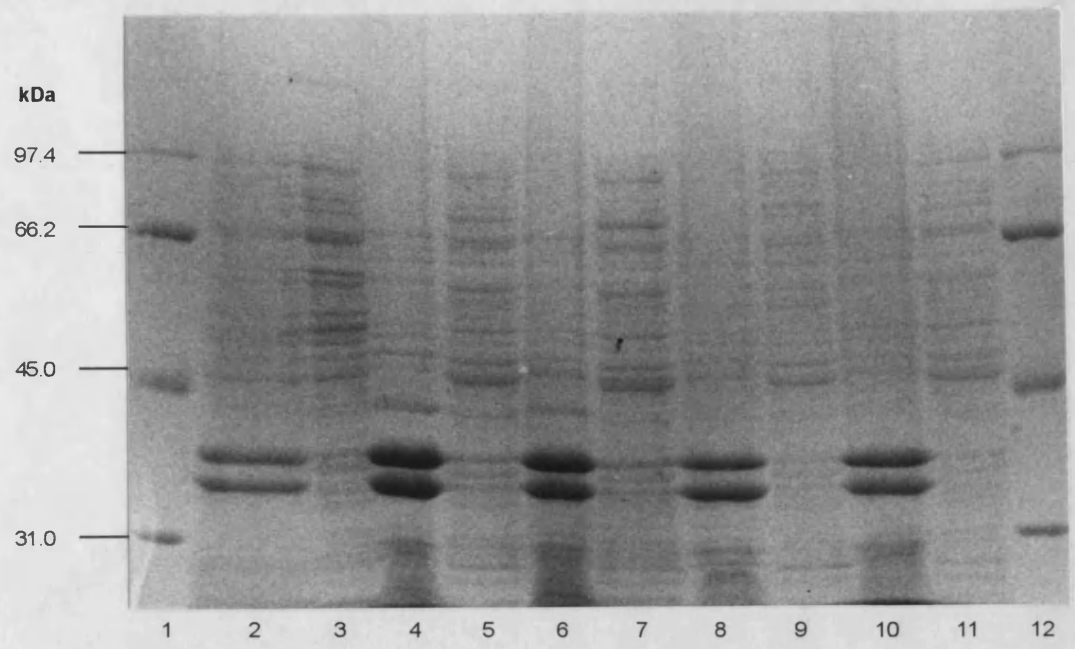


Fig. 6.4 **SDS-PAGE analysis illustrating the effects of sorbitol and betaine upon expression of soluble protein.**

Lanes 1 & 12: Protein standards, with the molecular weights indicated

Lanes 2 & 3: Cell debris (CD) & cell extract (CE), respectively, of *E. coli* strain DH5α, grown at 37°C in the presence of 5 mM betaine and 0.5 M sorbitol

Lanes 4 & 5: CD & CE of *E. coli* strain DH5α carrying the pKK223-3 plasmid-encoded truncated citrate synthase, grown in the presence of 5.0 mM betaine and 0.5 M sorbitol, at 30°C (for 20 h)

Lanes 6 & 7: As for lanes 4 & 5, but grown at 37°C (for 17 h)

Lanes 8 & 9: As for lanes 4 & 5, but grown in the presence of 5 mM betaine and 1.0 M sorbitol (for 30 h at 30°C)

Lanes 10 & 11: As for lanes 8 & 9, but grown at 37°C (for 24 h)

6.3 DISCUSSION

A truncated mutant, that involved removing the first 36 N-terminal residues of pig citrate synthase, was homology-modelled (section 6.2.1) and created (section 6.2.2). The purpose of this mutant was to investigate the role of the N-terminal arm upon the activity and thermostability of citrate synthase, since all archaeal citrate synthase sequences are of the shorter form and do not possess these N-terminal residues (Fig. 3.1). Although the arm could be related to phylogenetic characteristics, the absence of the residues in the *Tp. acidophilum* protein does mean that the thermophilic enzyme is a more compact and a less flexible structure when compared to pig citrate synthase. In the latter structure, the N-terminal residues constitute one of the four areas of the molecule that possesses the highest temperature factors [Remington *et al.*, 1982] and so one can imagine that as the temperature is increased this portion of the protein will start to unfold rapidly and possibly initiate denaturation within the protein. There are other examples where thermostable proteins have reduced flexibility at the termini when compared to their mesophilic equivalents. For example, with phosphoglycerate kinase from *B. stearothermophilus*, the N- and C-termini are both fixed to the main body of the protein with additional salt bridges [Davies *et al.*, 1993], as is the C-terminus in thermolysin from *B. thermoproteolyticus* [Pauptit *et al.*, 1988], and the N-terminus of rubredoxin from *P. furiosus* is fixed with additional hydrogen bonds [Blake *et al.*, 1991]. On the other hand, residues 22-39 of pig citrate synthase are involved in an extensive range of intra- and intersubunit contacts (section 6.2.1 and Remington *et al.* [1982]) and the loss of these hydrogen bonds and hydrophobic interactions could have a destabilising effect upon the structure of the mesophilic enzyme; however, analysis of the model of the mutant did not indicate any thermodynamic instability within the protein.

As far as expression of the mutant pig citrate synthase is concerned, both soluble and insoluble forms of the protein were produced (sections 6.2.3 and 6.2.4), but both were inactive and so other expression systems need to be tested in order to obtain an active and soluble enzyme. These studies could utilise the use of sorbitol and betaine in the growth medium, since it was under these conditions that the

soluble mutant protein was expressed (section 6.2.4). The aggregation of the truncated citrate synthase molecules could be the result of hydrophobic clumping initiated between the now-exposed residues of the protein, or the detrimental effect of removing the N-terminus of the protein so that correct folding upon protein biosynthesis is hindered. The guanidine.HCl treatment of the aggregated mutant enzyme was unsuccessful at refolding the protein into an active and soluble form, this may be rectified if the protocol was optimised using dialysis, for example, instead of dilution to remove the guanidine.HCl, or even using urea or NaCl in place of the guanidine.HCl. Refolding studies were not carried out on the soluble mutant protein, expressed in the presence of sorbitol and betaine (section 6.2.4), because the endogenous *E. coli* proteins could interfere with the refolding process; nevertheless, the experiment could be carried out on a purified sample of the soluble truncated protein. However, it does seem more likely that the truncated peptide is unable to fold into a structure with a global free energy minimum that represents the same structure as wild-type pig citrate synthase but without the N-terminal 36 residues.

The mutations carried out in the creation of the truncated pig citrate synthase were more dramatic than those of previous site-directed mutants, described in Chapters 4 and 5, where only one or two residues were altered. For this reason, the results obtained with expression of the truncated mutant are not altogether surprising, considering that expression of the cavity-filling mutants yielded inactive citrate synthase in both soluble and insoluble forms (sections 5.2.3 and 5.2.4). Once again the gap between theory and practice of site-directed mutagenesis has been highlighted, although there was more indication from the homology modelling in this chapter that a truncated mutant of citrate synthase was likely to be less stable than the cavity mutants, due to the loss of a significant number of hydrogen bonds.

If a soluble and active form of the truncated mutant of pig citrate synthase is produced, then the effect of the N-terminal arm upon the stability and activity of the enzyme can be characterised and it will be revealed if the N-terminus is simply a phylogenetic marker or if it actually plays a role in the thermostability of citrate synthase.

CHAPTER 7

PERSPECTIVES

The work described in this thesis, investigating protein thermostability in citrate synthase from the Archaea, complements that of our research group. A number of enzymes, in particular citrate synthase, are used as models for the study of protein thermal stability, halophilicity and, more recently, psychrophily. The genes for these enzymes have been cloned from archaeal species, sequenced and over-expressed, either in the bacterium *E. coli* or back in the modified archaeal strain. Over production of the protein of interest has then enabled its purification, characterisation and in most cases crystallisation. Using these 3D structures in comparative analyses with those of mesophilic homologs, we have gained an insight into possible features that could relate to the enhanced stability of the archaeal enzymes. The next stage of the research entails using site-directed mutagenesis to investigate the role of these features in protein stability. A full characterisation of the mutants, including structure elucidation, then determines whether or not the feature does play a role in stability.

The crystal structures of citrate synthase from pig, *Tp. acidophilum* and *P. furiosus* are available [Remington *et al.*, 1982; Russell *et al.*, 1994; Russell *et al.*, in press]. Comparative analyses between these three proteins have been conducted and have resulted in the proposal of a number of features that may play a role in conferring (hyper)thermostability upon the archaeal enzymes [described in Chapter 3 and Russell, 1994; Russell *et al.*, 1994; Russell *et al.*, in press]. The trend from mesophile to hyperthermophile, appears to be towards a more compact and potentially less flexible molecule, achieved in the thermostable proteins through additional interactions, fewer cavities, increased α-helix stability, better complementarity between the subunits of the dimer and an absence of flexible regions, such as loops and surface helices. A few of these features have been tested by site-directed mutagenesis, mutating pig citrate synthase to possess the feature observed in the *Tp. acidophilum* enzyme.

Creating mutations at the subunit interface (mutants S139A and S139A/T146A) seemed to destabilise the protein slightly, probably through small conformational changes in the secondary and tertiary structure (Chapter 4). On the other hand, the cavity-filling mutants (S412Y and A118Y/S412Y) and the truncated mutant of pig citrate synthase were all expressed in inactive forms (Chapters 5 and 6). The mutations were homology-modelled into the pig citrate synthase structure prior to performing the site-directed mutagenesis of the gene, and in each the case the results indicated that such the mutant protein would be stable. The expression of soluble and active protein with the truncated mutant may have been a naive expectation, considering that the removal of the N-terminus of the protein may have affected the process of protein folding upon biosynthesis and also resulted in the exposure of hydrophobic residues that could lead to aggregation of the protein molecules. However, the similar results obtained with the expression of the two cavity-filling mutants in (Chapter 5) were somewhat surprising and rather disappointing. These examples with the various mutants of pig citrate synthase have highlighted how the theory of homology-modelling used in this instance is only concerned with the conformational stability of the folded protein structure, and has not taken into account the process of protein folding upon synthesis of the polypeptide chain. Thus, the modelled structure of a potential mutant protein should be analysed with caution. All being well, if the mutant protein is expressed in a soluble form and can be crystallised then the effects of the mutation can be studied in the true crystal structure of the protein.

The results of the subunit interface site-directed mutagenesis studies have emphasised what little effect these single-site amino acid exchanges have upon the thermostability of the pig citrate synthase. In order to achieve more significant changes in the relative thermal stabilities of wild-type and mutant citrate synthases, the step forward may be to create proteins that incorporate larger mutations, for example swapping the domains between the mesophilic and thermophilic enzymes. To explain, the monomer of pig and *Tp. acidophilum* citrate synthase consists of two domains, the large domain which is involved in intersubunit contacts in the dimeric molecule and the small domain which forms part of the active site in the enzyme. The conservation and orientation of binding and catalytic residues within the active

site of citrate synthases from pig and *Tp. acidophilum* are in very close agreement. Thus, assuming that the domain-swapping mutant enzymes would retain activity, it would be interesting to find out what effect domain composition should have upon the relative thermostabilities of the two citrate synthase species.

In addition to characterising thermophily, members of our group are currently investigating the exact opposite, psychrophily. Will it be revealed that the psychrophiles have adapted to cooler habitats with the possession of proteins that have an increased flexibility and decreased compactness to their structure, i.e. features opposite to those observed in the two thermostable archaeal citrate synthases ?

Comparative analyses such as those mentioned above, highlight features that may be global to all thermostable proteins, although at the current stage in this research it appears that individual proteins have achieved the same goal of thermostability employing a variety of mechanisms. In addition to adding to the traffic rules, the analyses may also identify stabilising characteristics that are phylogenetically unique to each of the three domains of organisms. This type of characterisation can be accomplished in two stages, (i) by the comparison of mesophilic and thermophilic protein homologs whose host organisms are within the same domain, in order to identify protein thermostabilising features and (ii) a comparison of these features with those identified in the other two domains. Unfortunately with citrate synthase, we do not have the crystal structure of the enzyme from a mesophilic Archaeon. As with all the protein models studied so far, it is this lack of 3D structural data that has hampered the distinction between characteristics that are global to thermostable proteins and those that are unique to say, the Archaea. It is also unfortunate that only a small percentage of the protein models include archaeal structural homologs, quite disconcertingly so since the only hyperthermostable organisms isolated so far have been Archaea and thus proteins from these organisms are needed in order to characterise protein hyperthermostability.

The justification of research into characterising protein thermostability is often questioned, notably since biotechnology companies, that may require thermostable

enzymes, are more likely to screen a range of thermophilic organisms in search of the desired enzyme with the suitable stability properties, rather than mutate a mesophilic equivalent of the required protein in order to produce a more stable and active homolog. This choice is primarily due the lack of rules that guarantee to introduce stability into any mesophilic protein. However, as the research into protein thermostability progresses, those general rules are likely to emerge and at the same time the techniques of protein engineering and our understanding of protein folding *in vivo* will also improve. There will still be a bottle-neck to the approach of engineering stability into an enzyme though, and that is the determination of the true 3D structure of the protein required for informative decisions concerning the potential mutations. However, given the current rate of progress with molecular modelling techniques, the time should come when close approximations to the true 3D structures of any protein can be made from the gene sequence alone. This statement is particularly relevant considering that gene sequences are solved at a much greater rate than protein structures are elucidated.

There are also those applications where the protein of required thermostability does not actually exist and thus there is no alternative but to engineer the enhanced stability into the existing homolog. An example of one such application is with the potential need for a hyperthermostable glucose dehydrogenase, a metabolic enzyme found in the moderately thermostable *Tp. acidophilum* and the extreme thermostable Archaeon *S. solfotaricus*, but unfortunately not discovered, as yet, in any hyperthermostable organisms such as *P. furiosus*.

There is still much research to be carried out before thermostability of any protein can be defined by the same set of criteria, and although the mutagenesis work on pig citrate synthase described in this thesis has not led to the identification of positive stabilising features in the enzyme, hopefully some of the other suggestions for investigation described in section 3.3.2 and discussed above, will help shed light on the enhanced stability of the archaeal citrate synthase molecules, and in turn add to the traffic rules for general protein stability.

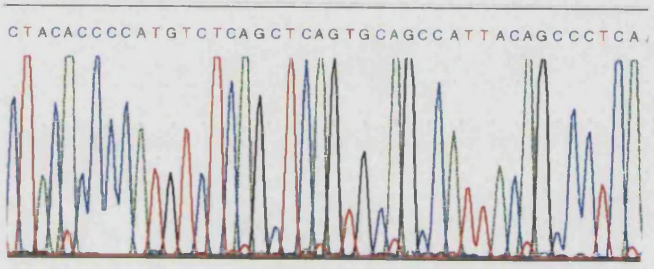
CHAPTER 8

APPENDIX

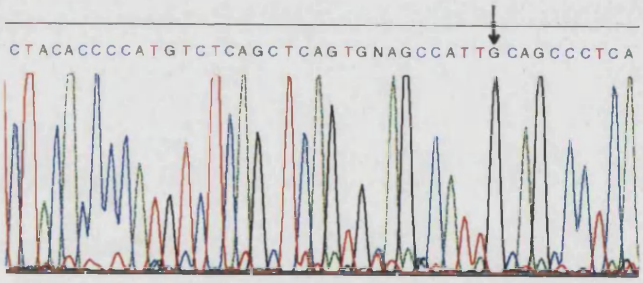
SEQUENCING RESULTS:

For the manual sequencing gels, the sequence reads from the bottom of the gel upwards and the lanes are labelled with their respective nucleotide. For the automated sequencing (carried out by J Bartley, University of Durham) double-stranded pKK223-3 vector carrying the respective genes was used. The chromatograms are shown with the sequence labelled above (reading right to left, and colour-coded).

(a) Wild-type



(b) T146A



(c) S139A/T146A

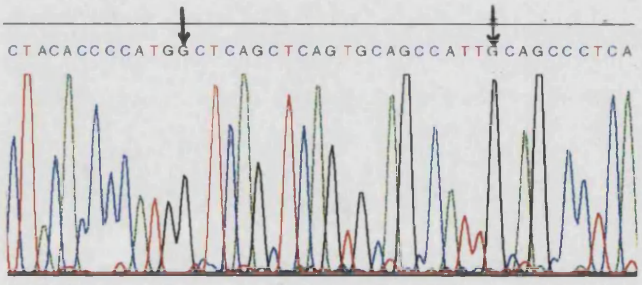


Fig. 1. (a) wild type pig citrate synthase
(b) the single interface mutant, T146A
(c) the double interface mutant, S139A/T146A. Bases 490 to 532 are shown. The arrows indicate the mutated nucleotides.

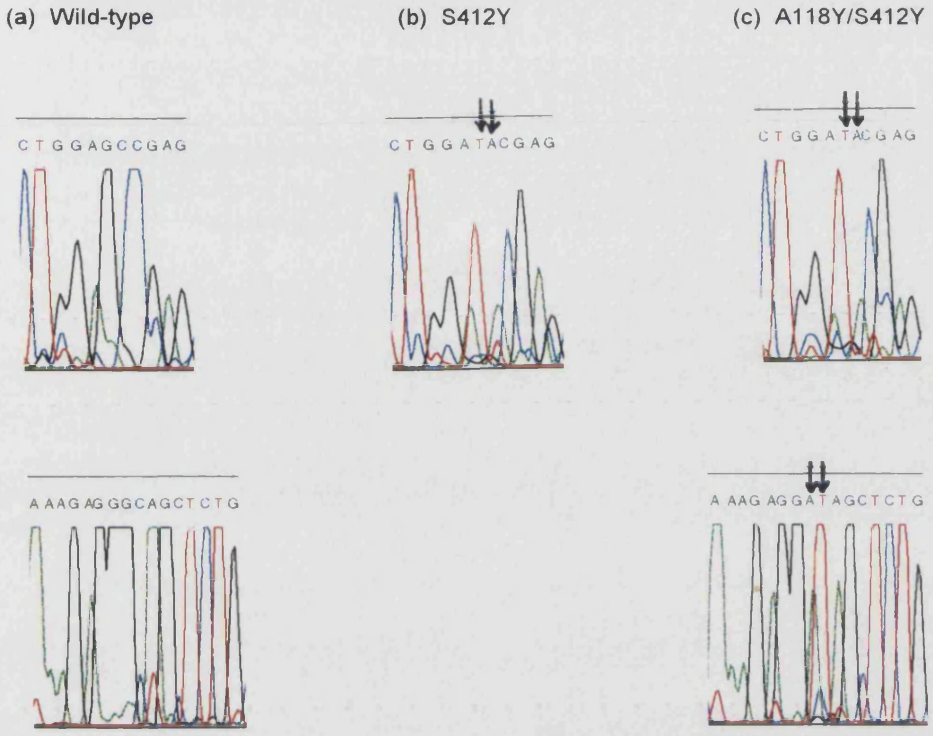


Fig. 2 (a) wild type pig citrate synthase
(b) the single cavity mutant, S412Y
(c) the double cavity mutant, A118Y/S412Y. Bases 432 to 447 (bottom row) and 1317 to 1327 (top row) are shown. The arrows indicate the mutated nucleotides.

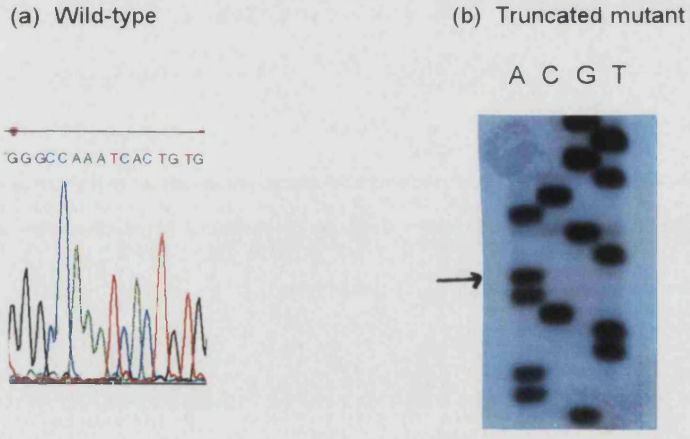


Fig. 3 (a) wild type pig citrate synthase (by automated sequencing)
(b) the truncated mutant of citrate synthase (by manual sequencing). Bases 186 to 201are shown. The arrow indicates the initiation codon for protein translation in the mutant.

CHAPTER 9

REFERENCES

Ahern, T. J. & Klibanov, A. M.
Science (1985) **228**, 1280-1284

Alber, T., Dao-Pin, S., Nye, J. A., Muchmore, D. C. & Matthews, B. W.
Biochemistry (1987) **26**, 3754-3758

Alter, G. M., Casazza, J. P., Zhi, W., Nemeth, P., Srere, P. A. & Evans, C.T.
Biochemistry (1990) **29**, 7557-7563

Amaki, Y., Nakano, H. & Yamane, T.
Appl. Microbiol. Biotech. (1994) **40,** 664-668

Anderson, D. E., Becktel, W. J. & Dahlquist, F. W.
Biochemistry (1990) **29**, 2403-2408

Anderson, S. C. K., Powrie, R. & Mitchell, C. G.
Biochem. Soc. Trans. (1993) **22**, 41S

Arakawa, T. & Timasheff, S. N.
Biophys. J. (1985) **47**, 411-414

Ashworth, J. M. & Kornberg, H. L.
Proc. R. Soc. Lond. Series B (1966) **165**, 179-188

Barns, S. M., Fundyga, R. E., Jeffries, M. W. & Pace, N. R.
Proc. Natl. Acad. Sci. USA (1991) **91**, 1609-1613

Barton, G. J. & Sternberg, M. J. E.
J. Mol. Biol. (1990) **212**, 389-402

Bell, J. A., Becktel, W. J., Sauer, U., Baase, W. A. & Matthews, B. W.
Biochemistry (1992) **31**, 3590-3596

Benachenhou-Lahfa, N., Forterre, P. & Labedan, B.
J. Mol. Evol. (1993) **36**, 335-346

Blake, P. R., Park, J. B., Bryant, F. O., Aono, S., Magnuson, J. K., Eccleston, E., Howard, J. B., Summers, M. F. & Adams, M. W. W.
Biochemistry (1991) **30**, 10885-10895

Blackwell, J. R. & Horgan, R.
FEBS Lett. (1991) **295**, 10-12

Bloxham, D.P., Parmelee, D.C., Kumar, S., Wade, R.D., Ericson, L.H., Neurath, H., Walsh, K.A. & Titani, K.
Proc. Natl. Acad. Sci. USA (1981) **78**, 5381-5385

Bloxham, D.P., Parmelee, D.C., Kumar, S., Walsh, K.A. & Titani, K.
Biochemistry (1982) **21**, 2028-2036

Böhm, G. & Jaenicke, R.
Prot. Eng. (1994) **7**, 213-220

Bowden, G. A. & Georgiou, G.
J. Biol. Chem. (1990) **265**, 16760-16766

Bowler, B. E., May, K., Zaragoza, T., York, P., Dong, A. & Caughey, W. S.
Biochemistry (1993) **32**, 183-190

Bradford, M. M.
Anal. Biochem. (1976) **72**, 248-254

Brems, D. N.
Biochemistry (1988) **27**, 4541-4546

Bright, J., Byrom, D., Danson, M. J., Hough, D. W. & Towner, P.
Eur. J. Biochem (1993) **180**, 549-554

Britton, K. L., Baker, P. J., Borges, K. M., Engel, P. C., Pasquo, A., Rice, D. W., Robb, F. T., Scandurra, R., Stillman, T. J. & Yip, K. S. P.
Eur. J. Biochem. (1995) **229**, 688-695

Brown, J. R. & Doolittle, W. F.
Proc. Natl. Acad. Sci. (1995) **92**, 2441-2445

Brunger, A. T.
Acta. Cryst. (1990) **A46**, 46-57

Buchner, J., Schmidt, M., Fuchs, M., Jaenicke, R., Rudolph, R., Schmid, F. X. & Kiefhaber, T.
Biochemistry (1991) **30**, 1586-1591

Burley, S. K. & Petsko, G. A.
Science (1985) **229**, 23-28

Chan, M. K., Makund, S., Kletzin, A., Adams, M. W. W. & Rees, D. C.
Science (1995) **267**, 1463-1469

Christiansen, C., Freundt, E. A. & Black, F. T.
Int. J. Syst. Bacteriol. (1975) **25**, 99-101

Cowan, D. A.
TIBTECH. (1992) **10**, 315-323

Daggett, V. & Levitt, M.
J. Mol. Biol. (1993) **232**, 600-619

Daniels, L.
The Biochemistry of the Archaea (Archaebacteria), (1993) **26**, 41-112, Elsevier Science Publishers B.V., Amsterdam

Danson, M. J.
Adv. Microbiol. Physiol. (1988) **29**, 165-230

Danson, M. J. & Hough, D. W.
Biochem. Soc. Symp. (1993) **58**, 7-21

Danson, M. J., Black, S. C., Woodland, L. L. & Wood, P. A.
FEBS Lett. (1985) **179**, 120-124

Danson, M.J., Harford, S. & Weitzman, P.D.J.
Eur. J. Biochem. (1979) **101**, 515-521

Danson, M. J., Hough, D. W. & Lunt, G. G.
The Archaebacteria: Biochemistry & Biotechnology, Biochemical Society Symposia (1992) **58**, Portland Press, London

Dao-pin, S., Sauer, U., Nicholson, H. & Matthews, B. W.
Biochemistry (1991) **30**, 7142-7153

Darland, G., Brock, T. D., Samsonoff, W. & Conti, S. F.
Science (1970) **170**, 1416-1418

David, M., Lubinsky-Mink, S., Ben-Zvi, A., Suissa, M. & Ulitzur, S.
Biochem. J. (1991) **278**, 225-234

Davies, G. J., Gamblin, S. J., Littlechild, J. A. & Watson, H. C.
Proteins: Structure, Function and Genetics (1993) **15**, 283-289

De Cordt, S., Hendrickx, M., Maesmans, G. & Tobback, P.
Biotech. Bioeng. (1994) **43**, 107-114

De Lange, R. J. & Williams, L. C.
J. Biol. Chem. (1981) **256**, 905-911

De Long, E. F., Wu, K. Y., Prézelin, B. B. & Jovine, R. V. M.
Nature (1994) **371**, 695-697

Devereux, J., Haeberli, P. & Smithies, O.
Nucleic Acids Res. (1984) **12**, 387-395

Dill, K. A.
Biochemistry (1990) **29**, 7133-7155

Donald, L.J. & Duckworth, H. W.
Biochem. Biophys. Res. Commun. (1986) **141**, 797-803

Donald, L.J., Molgat, G.F. & Duckworth, H.W.
J. Bacteriol. (1989) **171**, 5542-5550

Eliopolous, E. E., Geddes, A. J., Brett, M. Pappin, D. J. C. & Findlay, J. B. C.
Int. J. Biol. Macromol. (1982) **4**, 263

Else, A.J., Danson, M.J. & Weitzman, P.D.J.
Biochem. J. (1988) **254**, 347-442

Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P. & Matthews, B. W.
Science (1992) **255**, 178-183

Evans, C. T., Owens, D. D., Sumegi, B., Kispal, G. & Srere, P.
Biochemistry (1988) **27**, 4680-4686

Fersht, A. R., Shi, J.-P., Knill-Jones, J., Lowe, D. M., Wilkinson, A. J., Blow, D. M., Brick, P., Carter, P., Waye, M. M. Y. & Winter, G.
Nature (1985) **314**, 235-238

Fischer, B., Sumner, T. & Goodenough, P.
Biotech. Bioeng. (1993) **41**, 3-13

Forterre, P., Benachenhou-Lahfa, N., Confalonieri, F., Duguet, M., Elie, C. & Labedan, B.
Biosystems (1993) **28**, 15-32

Fox, G. E., Luehrsen, K. R. & Woese, C. R.
Zbl. Bakt. Hyg. (1982) **3**, 330-345

Fuhrman, J. A., McCallum, K. & Davis A. A.
Nature (1992) **356**, 148-149

Fukaya, M., Takemura, H., Okumara, H., Kawamura, Y., Horinouchi, S. & Beppu, T.
J. Bacteriol. (1990) **172**, 2097-2104

Gambacorta, A., Trincone, A., Nicolause, B., Lama, L. & De Rosa, M.
Syst. Appl. Microbiol. (1994) **16**, 518-527

Geiger, T. & Clarke, S.
J. Biol. Chem. (1987) **262**, 785-794

Gogarten, J. P., Kibak, H., Taiz, L., Bowman, E. J., Bowman, B. J., Manoison, M. P., Poole, R. J., Date, T., Oshima, T., Konishi, J., Dena, K. & Yoshida, M.
Proc. Natl. Acad. Sci. (1989) **86**, 6661-6665

Gokhale, R. S., Agarwalla, S., Francis, V. S., Santi, D. V. & Balaram, P.
J. Mol. Biol. (1994) **235**, 89-94

Golding, G. B. & Gupta, R. S.
Mol. Biol. Evol. (1995) **12**, 1-6

Gottschal, J. C. & Prins, R. A.
TREE (1991) **6**, 157-162

Grossebuter, W. & Gorisch, H.
Syst. Appl. Microbiol. (1985) **6**, 119-124

Gupta, R. S. & Singh, B.
Curr. Biology (1994) **4**, 1104-1114

Gutierrez, C., Barondess, J., Maniol, C. & Beckwith, J.
J. Mol. Biol. (1987) **195**, 289-297

Hanahan, D.
DNA Cloning, (1985) Volume 1 - a practical approach, ed. Glover, D. M., 109-135
[IRL Press, Oxford]

Harper, E. T. & Rose, G. D.
Biochemistry (1993) **32**, 7605-7609

Heinzen, R.A., Frazier, M.E. & Mallavia, L.P.
Gene (1991) **109**, 63-69

Hensel, R. & Jakob, I.
Syst. Appl. Microbiol. (1994) **16**, 742-745

Hol, W. C. J.
Prog. Biophys. Molec. Biol. (1985) **45**, 149-195

Horovitz, A., Matthews, J. M. & Fersht, A. R.
J. Mol. Biol. (1992) **227**, 560-568

Hough, D. W. & Danson, M. J.
Lett. Appl. Microbiol. (1989) **9**, 33-39

Hubbard, S. J., Gross, K.-H. & Argos, P.
Prot. Eng. (1994) **7**, 613-626

Ikai, A.
J. Biochem. (1980) **88**, 1895-1898

Imada, K., Sato, M., Tanaka, N., Katsube, Y., Matsuura, Y. & Oshima, T.
J. Mol. Biol. (1991) **222**, 725-738

Ishikawa, K., Kimura, S., Kanaya, S., Morikawa, K. & Nakamura, H.
Prot. Eng. (1993(c)) **6**, 85-91

Ishikawa, K., Nakamura, H., Morikawa, K. & Kanaya, S.
Biochemistry (1993(b)) **32**, 6171-6178

Ishikawa, K., Okumura, M., Katayanagi, K., Kimura, S., Kanaya, S., Nakamura, H. & Morikawa, K.
J. Mol. Biol. (1993(a)) **230**, 529-542

Iwabe, N., Kuma, K.-I., Hasegawa, M., Osawa, S. & Miyata, T.
Proc. Natl. Acad. Sci. USA (1989) **86**, 9355-9359

Jaenicke, R.
Eur. J. Biochem (1991) **202**, 715-728

Jaenicke, R.
Prog. Biophys. Mol. Biol. (1987) **49**, 117-237

Jaenicke, R. & Rudolph, R.
Protein Structure: A Practical Approach, ed. Creighton, T. E. (1989) 191-223, [IRL Press, Oxford]

James, K. D., Bonete, M. J., Byrom, D., Danson, M. J. & Hough, D. W.
Biochem. Soc. Trans. (1991) **20**, 12S

Jin, S. & Sonenshein, A. L.
J. Bacteriol. (1994) **176**, 4669-4679

John, J. J., Crennell, S. J., Hough, D. W., Danson, M. J. & Taylor, G. L.
Structure (1994) **2**, 385-393

Johansson, C. J. & Pettersson, G.
Eur. J. Biochem. (1974) **42**, 383-388

Johansson, C. J. & Pettersson, G.
Biochem. Biophys. Acta. (1977) **484**, 208-215

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M.
Acta. Cryst. (1991) **A47**, 110-119

Kabsch, W. & Sander, C.
Biopolymers (1983) **22**, 2577-2637

Kallwass, H. K. W., Surewicz, W. K., Parins, W., MacFarlane, C. L. A., Luyten, M. A., Kay, C. M., Gold, M. & Jones, J. B.
Prot. Eng. (1992) **5**, 769-774

Karpusas, M., Baase, W. A., Matsumara, M. & Matthews, B. W.
Proc. Natl. Acad. Sci. USA (1989) **86**, 8237-8241

Karpusas, M., Branchaud, B. & Remington, S.J.
Biochemistry (1990) **29**, 2213-2219

Karpusas, M., Holland, D. & Remington, S. J.
Biochemistry (1991) **30**, 6024-6031

Kates, M., Kushner, D. J. & Matheson, T. A.
The Biochemistry of the Archaea (Archaebacteria), (1993) **26**, Elsevier Science Publishers B.V., Amsterdam

Kellis, J. T., Nyberg, K., Sali, D. & Fersht A. R.
Nature (1988) **333**, 784-786

Kelly, C. A., Nishiyama, M., Ohnishi, Y., Beppu, T. & Birktoft, J. J.
Biochemistry (1993) **32**, 3913-3922

Kirino, H., Aoki, M., Aoshima, M., Hayashi, Y., Ohba, M., Yamagishi, A., Wakagi, T. & Oshima, T.
Eur. J. Biochem. (1994) **220**, 275-281

Klenk, H.-P., Renner, O., Schwass, V. & Zillig, W.
Nucleic Acids Res. (1992) **20**, 5226

Kleywegt, G. A. & Jones, T. A.
Acta. Cryst. (1994), **D50**, 178

Komdörfer, I., Steipe, B., Huber, R., Tomschy, A. & Jaenicke, R.
J. Mol. Biol. (1995) **246**, 511-521

Kotik, M. & Zuber, H.
Eur. J. Biochem. (1993) **211**, 267-280

Krebs, H.A. & Lowenstein, J.M.
'Metabolic Pathways' (1960) 2nd ed., vol 1, pp. 129-203

Laemmli, U. K.
Nature (1970) **227**, 680-685

Lake, J. A.
TIBS (1991) **16**, 46-50

Laskowski, R. A., MacArthur, M. W., Moss, D. W. & Thornton, D. W.
J. Appl. Cryst. (1993) **26**, 283-291

LaVallie, E. R., DiBlasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F. & McCoy, J. M.
BIO/TECH (1993) **11**, 187-193

Lill, U., Lefrank, S., Henschen, A. & Eggerer, H.
Eur. J. Biochem. (1992) **208**, 459-466

Lohlein-Werhahn, G., Goepfert, P. & Eggerer, H.
Biol. Chem. Hoppe-Seyler (1988) **369**, 109-113

Luthy, R., Bowie, J. U. & Eisenberg, D.
Nature (1992) **356**, 83-85

Makhatadze, G. I. & Privalov, P. L.
J. Mol. Biol. **213**, 375-384

Matsumura, M., Becktel, W. J., Levitt, M. & Matthews, B. W.
Proc. Natl. Acad. Sci. USA (1989) **86**, 6562-6566

Matthews, B. W., Nicholson, H. & Becktel, W. J.
Proc. Nat. Acad. Sci. (1987) **84**, 6663-6667

McEvily, A. J. & Harrison, J. H.
J. Biol. Chem. (1986) **261**, 2593-2598

Menendez-Arias, L. & Argos, P.
J. Mol. Biol. (1989) **206**, 397-406

Muir, J. M., Russell, R. J. M., Hough, D. W. & Danson, M. J.
Prot. Eng. (1995) **8**, 583-592

Ner, S.S., Bhayana, V., Bell, A.W., Giles, I.G., Duckworth, H.W. & Bloxham, D.P.
Biochemistry (1983) **22**, 5243-5249

Nicholls, A., Sharp, K. A. & Honig, B.
Proteins: Structure, Function and Genetics (1991) **11**, 281

Numata, O., Takemasa, T., Takagi, I., Hirono, M., Hirano, H., Chiba, J. & Watanabe, Y.
Biochem. Biophys. Res. Commun. (1991) **174**, 1028-1034

Olsen, G. J. & Woese, C. R.
Can. J. Microbiol. (1989) **35**, 119-123

O'Neil, K. T. &DeGrado, W. F.
Science (1990) **250**, 646-651

Pace, C. N.
TIBS (1990) **15**, 14-17

Pace, C. N.
J. Mol. Biol. (1992) **226**, 29-35

Pakula, A. A. & Sauer, R. T.
Ann. Rev. Genet. (1989) **23**, 289-310

Patton, A. J., Hough, D. W., Towner, P. & Danson, M. J.
Eur. J. Biochem. (1993) **214**, 75-81

Pauptit, R. A, Karlsson, R., Picot, D., Jenkins, J. A., Niklaus-Reimer, A. S. & Jansonius, J. N.
J. Mol. Biol. (1988) **199**, 525-537

Pechmann, H., Tesch, A. & Klink, F.
FEMS Microbiol. Lett. (1991) **79**, 51-56

Perutz, M. F. & Raidt, H.
Nature (1975) **255**, 256-259

Privalov, P. L. & Makhatadze, G. I.
J. Mol. Biol. **213**, 385-391

Provencher, S. W. & Glöckner, J.
Biochemistry (1981) **20,** 33-37

Ptitsyn, O. B. & Finkelstein, A. V.
Biopolymers (1983) **22**, 15-25

Qaw, F. S. & Brewer, J. M.
Mol. Cell. Biochem. (1986) **71**, 121-127

Ree, H. K., Larsen, N., Gutell, R. G. & Zimmermann, R. A.
Syst. Appl. Microbiol. (1993) **16**, 333-341

Regan, L. & Degrado, W. F.
Science (1988) **241**, 976-978

Remington, S. J.
Curr. Top. Cell. Regul. (1992) **33**, 209-219

Remington, S.J, Wiegand, G. & Huber, R.
J. Mol. Biol. (1982) **158**, 111-152

Rennell, D., Bouvier, S. E., Hardy, L. W. & Poteete, A. R.
J. Mol. Biol. (1991) **222**, 67-87

Rentier-Delrue, F., Mande, S. C., Moyens, S., Terpstra, P., Mainfroid, V., Goraj, K., Lion, M., Hol, W. G. J. & Martial, J. A.
J. Mol. Biol. (1993) **229**, 85-93

Rosenkrantz, M., Alam, T., Kim, K. S., Clark, B. J., Srere, P. A. & Guarente, L. P.
Mol. Cell. Biol. (1986) **6**, 4509-4515

Rozema, D. & Gellman, S. H.
J. Am. Chem. Soc. (1995) **117**, 2373-2374

Rubin. B.H., Stallings, W. C. Glusker, J. P. Bayer, M. E., Janin, J. & Srere, P. A.
J. Biol. Chem. (1983) **258**, 1297-1298

Russell, R. J. M.
PhD Thesis, University of Bath, 1994

Russell, R. J. M., Hough, D. W., Danson, M. J. & Taylor, G. L.
Structure (1994) **2**, 1157-1167

Šali, D., Bycroft, M. & Fersht, A. R.
J. Mol. Biol. (1985) **220**, 779-788

Sambrook, J., Fritsch, E. F. & Maniatis, T.
'Molecular Cloning - A laboratory manual', Cold Spring Harbor Laboratory Press, Cold Spring Laboratory, New York

Sanger, F., Nicklen, S. & Coulson, A. R.
Proc. Natl. Acad. Sci. USA (1977) **74**, 5463-5467

Schendel, F.J., August, P.R., Anderson, C.R., Hanson, R.S. & Flickinger, M.C.
Appl. Environ. Microbiol. (1992) **58**, 335-345

Scholtz, S., Sonnenbichler, J., Schäfer, W. & Hensel, R.
FEBS Lett. (1992) **306**, 239-242

Schönheit, P.
The Biochemistry of the Archaea (Archaebacteria), (1993) **26**, 113-172, Elsevier Science Publishers B.V., Amsterdam

Schein, C. H. & Noteborn, M. H. M.
BIO/TECH (1988) **6**, 291-294

Searcy, D. G.
Biochem. Biophys. Acta. (1975) **395**, 535-547

Searcy, D. G.
Syst. Appl. Microbiol. (1986) **7**, 198-201

Segerer, A., Langworthy, T. A. & Stetter, K. O.
Syst. Appl. Microbiol. (1988) **10**, 161-171

Serrano, L., Horovitz, A., Avron, B., Bycroft, M. & Fersht, A. R.
Biochemistry (1990) **29**, 9343-9352

Serrano, L., Kellis, J. T., Cann, P., Mataouschek, A. & Fersht, A. R.
J. Mol. Biol. (1992(a)) **224**, 783-804

Serrano, L., Neira, J. L., Sancho, J. & Fersht, A. R.
Nature (1992(b)) **356**, 453-455

Serrano, L., Sancho, J., Hirshberg, M. & Fersht, A. R.
J. Mol. Biol. (1992(c)) **227**, 544-559

Shirley, B. A., Stanssens, P., Hahn, U. & Pace, C. N.
Biochemistry (1992) **31**, 725-732

Shortle, D., Stites, W. E. & Meeker, A. K.
Biochemistry (1990) **29**, 8033-8041

Smith, L. D., Stevenson, K. J., Hough, D. W. & Danson, M. J.
FEBS Lett. (1987) **225**, 277-281

Srere, P. A., Brazil, H. & Gonen, L.
Acta. Chem. Scand. (1963) **17**, S129-S134

Stein, D. B. & Searcy, D. G.
Science (1978) **202**, 219-221

Stephenson, R. C. & Clarke, S.
J. Biol. Chem. (1989) **264**, 6164-6170

Stuart, D. L.
PhD Thesis, University of Bristol, 1979

Stuart, D. L., Levine, M., Muirhead, H. & Stammers, D. K.
J. Mol. Biol. (1979) **134**, 109-142

Suissa, M., Suda, K. & Schatz, G.
EMBO J. (1984) **3**, 1773-1781

Sutherland, K.J., Henneke, C.M., Towner, P., Hough, D.W. & Danson, M.J.
Eur. J. Biochem. (1990) **194**, 839-844

Tesch, A. & Klink, F.
FEMS Microbiol. Lett. (1990) **71**, 293-298

Tomschy, A., Böhm, G. & Jaenicke, R.
Prot. Eng. (1994) **7**, 1471-1478

Unger, E. A., Hand, J. M., Cashmore, A. R. & Vasconcelos, A. C.
Plant Mol. Biol. (1989) **13**, 411-418

Wakabayashi, S., Fujimoto, N., Wada, K., Matsubara, H., Kersher, L. & Oesterhelt, D.
FEBS Lett. (1983) **162**, 21-24

Walker, J. E., Wonacott, A. J. & Harris, J. I.
Eur. J. Biochem. (1980) **108**, 581-586

Weitzman, P.D.J.
Adv. Microb. Physiol. (1981) **22**, 185-244

Weitzman, P.D.J. & Danson, M.J.
Curr. Topics Cell Regul. (1976) **10**, 161-204

Woese, C. R.
Microbiol. Rev. (1987) **51**, 221-271

Woese, C. R. & Fox, G. E.
Proc. Natl. Acad. Sci. USA (1977) **74**, 5088-5090

Woese, C. R. & Olsen, G. J.
Syst. Appl. Microbiol. (1986) **7**, 161-177

Woese, C. R. & Wolfe, R. S. (eds)
'The Bacteria' (1985) vol. 8, Academic Press, London

Woese, C. R., Kandler, O. & Wheelis, M. L.
Proc. Natl. Acad. Sci. USA (1990) **87**, 4576-4579

Woese, C. R., Magrum, L. J. & Fox, G. E.
J. Mol. Evol. (1978) **11**, 243-252

Wolters, J. & Erdmann, V. A.
J. Mol. Evol. (1986) **24**, 152-166

Wood, D.O., Atkinson, W.H., Sikorski, R.S. & Winkler, H.
J. Bacteriol. (1987) **169**, 3564-3572

Wu, J.-Y. & Yang, J. T.
J. Biol. Chem. (1970) **245**, 212-218

Yang, D., Kaine, B. P. & Woese, C. R.
Syst. Appl. Microbiol. (1985) **6**, 251-256

Zale, S. E. & Klibanov, A. M.
Biochemistry (1991) **25**, 5432-5444

Zellner, G. & Kneifel, H.
Arch. Microbiol. (1993) **159**, 472-476

Zhi, W., Srere, P. A. & Evans, C. T.
Biochemistry (1991) **30**, 9281-9286

Zillig, W., Schnabel, R. & Stetter, K. O.
Curr. Top. Microbiol. Immunol. (1985) **114**, 1-18

Zillig, W., Klenk, H. P., Palm, P., Leffers, H., Puhler, G., Gropp, F. & Garrett, R.
Endocytobiosis Cell. Res. (1989) **6**, 1-25